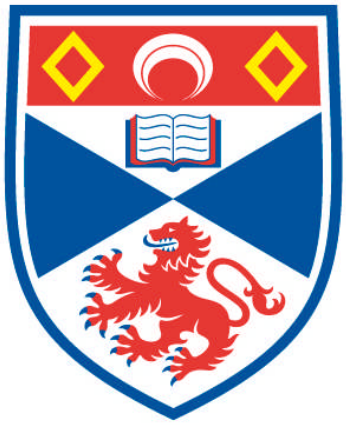
INFLUENCE OF NATURAL FATORS AND ANTHROPOGENIC STRESSORS ON SPERM WHALE FORAGING EFFORT AND SUCCESS AT HIGH LATITUDES

Saana Isojunno

A Thesis Submitted for the Degree of PhD
at the
University of St Andrews



2015

INFLUENCE OF NATURAL FACTORS AND ANTHROPOGENIC STRESSORS ON SPERM WHALE FORAGING EFFORT AND SUCCESS AT HIGH LATITUDES

Saana Isojunno

This thesis is submitted in partial fulfilment for the degree of PhD

at the University of St Andrews

July 2014

Abstract

Behavioural responses can reveal important fitness trade-offs and ecological traps in evolutionarily novel contexts created by anthropogenic stimuli, and are of increasing conservation concern due to possible links to population-level impacts. This thesis illustrates the use of proxies for energy acquisition and expenditure within multivariate and state-based modelling approaches to quantify the relative time and energetic costs of behavioural disturbance for a deep-diving marine mammal (*Physeter macrocephalus*) in foraging grounds in Kaikoura Canyon (New Zealand) and near Lofoten Islands (Norway). A conceptual framework is first developed to identify and explore links between individual motivation, condition and external constraints to behavioural disturbance [Chapter 1]. The following chapters then use data from behavioural response studies (BRS) to: 1) derive biologically relevant metrics of behaviour [all chapters], 2) investigate effects of boat-based focal follows and tagging procedures [Chapters 2-3], and 3) relate responses to specific disturbance stimuli (distance, approach, noise) from whale-watching [Chapter 2], naval sonar and playback of presumed natural predator (killer whale *Orcinus orca*) sounds [Chapter 4]. A novel hidden state model was developed to estimate behavioural budgets of tagged sperm whales from multiple streams of biologging (DTAG) data [Chapter 3]. Sperm whales traded off time spent at foraging depths in a non-foraging and non-resting state in response to both tag boat presence, 1-2 kHz naval sonar (SPL 131-165 rms re 1µPa) and mammal-eating killer whale sound playbacks, indicating that parallel non-lethal costs were incurred in both anthropogenic disturbance and presumed antipredatory contexts. While behavioural responses were highly variable by individual, biologically informed state-based models appeared effective to control for variability in energy proxies across different functional contexts. These results and Chapter 5 “linking buzzes to prey” demonstrate that behavioural context is a signal that can aid understanding of how individual non-lethal disturbance responses can impact fitness.

I, Saana Isojunno, hereby certify that this thesis, which is approximately 60 000 words in length, has been written by me, and that it is the record of work carried out by me, or principally by myself in collaboration with others as acknowledged, and that it has not been submitted in any previous application for a higher degree.

I was admitted as a research student in June, 2009 and as a candidate for the degree of Ph.D. in September, 2009; the higher study for which this is a record was carried out in the University of St Andrews between 2009 and 2014.

Date …… Signature of candidate ……

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of Ph.D in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree.

Date …… Signature of supervisor ……

Permission for publication

In submitting this thesis to the University of St Andrews I understand that I am giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. I also understand that the title and the abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker, that my thesis will be electronically accessible for personal or research use, and that the library has the right to migrate my thesis into new electronic forms as required to ensure continued access to the thesis. I have obtained any third-party copyright permissions that may be required in order to allow such access and migration.

Date …… Signature of candidate ……

Date …… Signature of supervisor ……

Statement of collaborations

I designed the visual and acoustic tracking protocol for Chapter 2 in collaboration with Jonathan Gordon and Manuel Fernandes. Manuel Fernandes and I established volunteer protocols and I was responsible for setting up all databases. While I stayed on the field for the initial five months in 2009-2010, Manuel Fernandes conducted the remaining field work and released the data for my use as they became available. Chapter 2 is partially published as a chapter in a report for the New Zealand department of Conservation. The document can be retrieved at http://doc.org.nz/documents/conservation/native-animals/marine-mammals/nelson-marlborough/Effects-of-tourism-on-sperm-whales.pdf.

Data for Chapters 3-5 were collected and supplied for me by the 3S (Sea mammals, Sonar, Safety) research project. Tagging data from 2005 were supplied by a collaborative study including Mark Johnson and Peter Madsen. I was not involved in the field work, and the controlled exposure experiments were fully designed by the 3S project. Received sound levels for the experimental exposures were estimated by Paul Wensween (Section 4.2.2).

Acknowledgments

I am indebted to my three supervisors Jonathan Gordon, Paul While and Patrick Miller. Jonathan for initiating me on sperm whales on the field and always reminding me of the biology behind numbers; Paul for in-depth education in acoustics and signal processing with enviable sense of humour; and Patrick Miller, for steering, focus and inspiration when I needed it most, and for being part of his prolific and incredibly supportive lab. The lab includes some serious talent and I am humbled to have become part of the gang: Kagari Aoki, Olga Filatova, Miguel Neves, Sara Tavares, Tomoko Narazaki, Filipa Samarra and Paul Wensveen. Thank you so much for supporting me all the way through!

A great many people were behind my Chapter 2 “effects of whale-watching”, and I would especially like to thank Manuel Fernandes for his enthusiasm and patience collecting data on the field together, and our continued collaboration since. I would also like to thank the principal investigators of the project Jonathan Gordon, Tim Markoviz and Christoph Richter and the managers and staff of Whale Watch Kaikoura, Wings Over Whales, Kaikoura Aeroclub and Kaikoura Helicopters for a phenomenal cross-platform collaboration. I’m so grateful for all our volunteers onboard Titi: Lorna Deppe, Anna Liljestrand, Hannah Curtis, Helen Fairlamb, Thomas barnfield, Heather Vance, John Hackett, Claire Louise Giner, Emma Ball, Jean Mathias Jourdier, Alicia Sanchez, Petrina Tay, Jessica Riggin, Leila Fouda, Isabelle Sikora, Eriko Fukuda, Bryony Manley, Valérie Lubrick, Kirsty Everley, Stéphanie Czudaj, James Miles, Cécile Houllé, Lindsey Nielsen, Adrien Lambrechts, Claire Hallier, Jason Gayton, Katie-Anne Isabella, Pete Weiner, Sarah Barry, Kerry Froud, Jessica Riggin, Courtnay Wilson, Jenipher Cate, Thomas Robinson, Kate Sprogis, Fergus Grant, Kalliopi Ghikopoulou, Benjamin Williams, Cally Meredith, Sue Jacson, Elena Trentin, Ruben Venegas, Gemma Veneruso, Sol Heber, Lise Sibilski, Rose Thorogood, Nasman, Chanyo Bunpunloot, Rémi Bigonneau, Dean Harliwich, Sandra Nussbaum, Jamie Macaulay, Sara Anderson, Daniel Lindner and Michele Gates. As well as work, Kaikoura was a special experience and I made life-long friends. I send the Kaikoura vibes to Popi, Kate, JM, Gemma and Elena especially.

Sincerest thanks to Mark Johnson, Peter Madsen, and all 3S (Sea mammals, Sonar, Safety) team members for both efforts on the field data collection in Norway and making the data available to this thesis. I’d especially like to acknowledge the 3S principal

investigators Petter Kvadsheim, Frans-Peter Lam and Peter Tyack for enabling the 3S collaboration along with their partners and funders (UK Ministry of Defence, U.S. Office of Naval Research, and World Wildlife Fund, Norway). I'm grateful for the 3S PI:s as well as the 3S P&P (PhD and post-doc) group, including Charlotte Curé, Paul Wensween, Lise Doksæter Sivle and Fleur Visser, who have all been fundamental to my learning and maturing ideas through discussion, presentations, and editorial comments.

For in-depth analytical support and advice throughout the hidden state model development process I also owe a big thanks to the team in Centre for Research into Ecological and Environmental Modelling (CREEM, St Andrews) including Len Thomas, Catriona Harris, Stacy DeRuiter, Dina Sadykova, Roland Langrock and Tiago Marques.

The sugar in the bottom of my acknowledgements is my Steve, the most patient and understanding friend throughout the thesis process - and husband to be tomorrow on the 12th of June!

Table of contents

1 GENERAL INTRODUCTION

1.1 What drives behavioural responses to disturbance stimuli?

Animal decision making is ultimately shaped by the benefits, costs and risks of behaviour that have been realised and selected upon in the past. Humans alter the environment within which species have evolved, and can pose evolutionarily novel situations that can elicit both adaptive and non-adaptive behaviours relative to the new selection pressures (Sih 2013). A particular concern is that even seemingly benign anthropogenic activities can elicit disturbance responses that trade off perceived reduction in risk against individual's time and energy investment on fitness enhancing activities; if persistent, such effects might lead to increased population vulnerability (Gill et al 2001, Frid and Dill 2002, Beale 2007, Lusseau and Bejder 2007, Bejder et al 2009, Berger-Tal et al 2011, Ellison et al 2011, Sih 2013). Such trade-offs are partly expected because animal species have evolved response and learning strategies to generalised threatening stimuli, and evaluate the cost of behavioural change (such as leaving a high quality patch) against perceived risks and rewards of a stimulus (Frid and Dill 2002, Brown et al 2004, Sih 2013). Prey species in particular can be expected to over-estimate risk when the costs of ignoring novel predation risks have been high in the past (Frid and Dill 2002). To mitigate and predict consequences of anthropogenic disturbance therefore, the challenge is to understand the adaptive significance of disturbance responses, quantify incurred costs as a function of perceived risk-rewards, and apply these costs across time-scales and contexts. Although lack of behavioural response does not necessarily imply lack of perceived threat and individual vulnerability (Gill et al 2001, Beale 2007, Ellison et al 2011), adaptive theory predicts that animals should only trade off fitness-enhancing activities in a given context when the anthropogenic stimulus is perceived as a risk or a cost.

Both proximate causes (individual state and environmental conditions) and ultimate (evolutionary) causes for behavioural responses are important to understand and quantify the consequences of behavioural disturbance. For example, a proximate cause in an anthropogenic noise study could be the sensation level, or ‘dose’ of an acoustic exposure: sound propagation, ambient noise levels in the environment (e.g. Dunlop et al 2013), receiver's hearing threshold at specific frequencies, and exposure in terms of the temporal characteristics of the signal (Madsen 2005, Nowacek et al 2007, Southall et al 2008, Ellison

et al 2011). Ultimately however, animals have imperfect knowledge of the risks and costs associated with such anthropogenic stimuli, as of their environment (e.g. predators, conspecifics, weather). “An animal should be a statistician” (Pyke 1984) in order to make optimal behavioural decisions in a variable environment. Individuals’ perception of a novel stimulus can be based upon both very specific 'ghosts of signals past' (Sih 2013), such as acoustic spectral content and fine-structure resembling a natural predator (e.g. Tyack et al 2011), and more universal features, such as size, number, speed and directness of approach that may help recognition of predators or signal predator intent (Frid and Dill 2002, Stankowich and Blumstein 2005). Anthropogenic signals can mimic such generalized stimuli unintentionally or intentionally (Frid and Dill 2002, Sih 2003, Stankowich and Blumstein 2005).

Signal perception and tolerance level to disturbance may also vary with behavioural and environmental context, internal state, and experience (Beale 2007, Southall et al 2008, Bejder et al 2009, Ellison et al 2011). An ability to modify perception of a threat through experience is likely to be an adaptive trait in many species, and may lead to habituation or sensitisation over time (Bejder et al 2009). Behavioural and environmental context, in turn, can determine whether a noise is perceived as a cost or nuisance to current activity (e.g. group size increases vigilance and dilution) and what other behavioural options are available, such as leaving habitat (Gill et al 2001, Lusseau and Bejder 2007). Behavioural options may be environmentally constrained, e.g. due to patch quality and connectivity, distance to refuge, effects on crypsis (Stankowich and Blumstein 2005) or effects on acoustic communication (Lombard effect, Hotchkin and Parks 2013). Predator density is also expected to affect responsiveness, with cumulatively higher cost of vigilance in higher density areas (Stankowich and Blumstein 2005). Individual’s age, physiological, energetic and reproductive state as well as behavioural syndromes can further influence the cost, benefit or even ability to respond or learn (e.g. shy vs. bold strategists, Dingemanse and Wolf 2010). Furthermore, costs can be incurred before or independently of a visible behavioural change – as part of a physiological stress response where the sympatric nervous system triggers the release of adrenalin and noradrenalin (Romero 2004, Wright et al 2011). Thus, the probability or intensity of a behavioural response does not necessarily have a

simple positive relationship with the vulnerability of an individual (Gill et al 2001, Beale 2007, Ellison et al 2011). The fitness cost of behavioural responses therefore represents a *minimum* cost incurred by disturbance (Beale 2007).

Detecting numeric (population) changes in distribution and abundance of sparsely distributed and highly mobile marine species can be challenging (Evans and Hammond 2004), costly and take years of monitoring to acquire sufficient statistical power to detect declines, notably in already small populations (e.g. vaquita, Taylor and Gerrodette 1993) and deep-diving marine mammals with short surface durations (e.g. beaked whales, Barlow and Gisinger 2006). Measuring acute responses in individual behaviour can be a more effective means to detect change and potentially inform about the causes of population change. For any such changes to be transferable to a population level however, the sampled behaviour must have a significant contribution to critical life functions, such as feeding or breeding (Population Consequences of Acoustic Disturbance (PCAD), NRA 2005). Many of the individual and contextual variables can be controlled or measured during controlled exposure experiments (CEE), but careful inference and complimentary observational data are needed to extrapolate short-term responses to long-term and real populations (Lusseau and Bejder 2007, Tyack 2009). In their "conservation behaviour framework", Berget-Tal et al (2011) defined three key behavioural domains that are of concern: 1) movement and space-use, 2) foraging and predator-prey relationships, and 3) social behaviour and reproduction. Changes in these behavioural domains may be used as early warning indicators of a numeric consequence and to monitor effectiveness of management programs (Berger-Tal 2011).

With advances in biotelemetry (Cooke et al 2004, Johnson et al 2009), a promising approach is to use tag sensor data to derive metrics that can serve as proxy indicators of the benefits and costs of behaviour, such as energy (e.g. proxies of active expenditure from tri-axis accelerometers, Wilson et al 2006), information (e.g. biosonar target range estimates, Madsen et al 2005b), conspecific associations (e.g. vocal responses indicate initiation of social response to predator presence, Curé et al 2013) or body condition (e.g. drift rates related to buoyancy, Biuw et al 2003). These proxies make an implicit assumption of the proximate or ultimate motivation of behaviours and are therefore more directly linked to biological consequences of behaviour. Typically, proxies can be quantified as a ‘currency’

measured over a study window or time windows that match the duration and spatial scale of the relevant behaviour (such as prey encounter rates during foraging phases, Watwood et al 2006). A more realistic approach is to allow time windows or behavioural states to vary over time (state-switching). States may be classified directly from data ('behavioural state'), such as area-restricted movement, or refer to an underlying motivation that drives observed behaviour ('motivational state', Bindra 1978), such as hunger level. Internal and external drivers of behaviour may be combined by considering behavioural time series to arise from discrete functional units ('functional states') that are associated with the fulfilment of a particular proximate or ultimate goal or set of goals based upon a priori hypotheses (Nathan et al 2008). Defining individual behaviour in terms of discrete states not only allows for explicit definition of state-specific currencies, but also any currency flow across states or a hierarchy of states (e.g. compensation or ecological carry-over effects, O'Connor et al 2014). With advances in statistical computing, there is increasing scope to estimate these states within more realistic hidden process models that distinguish the observation and underlying ('hidden') process explicitly (Patterson et al 2008, Schick et al 2008). For movement, this process might comprise of a navigation capacity (where to move) and motion capacity (how to move) interacting with individual's internal (physiological and psychological) state and external environment (Getz and Saltz 2008, Nathan et al 2008).

1.2 A conceptual model

Here, a framework is proposed where the cost of a behavioural response is partially controlled by the motivational state of the individual. This framework extends the movement ecology paradigm by Nathan et al (2008) to incorporate proxy currencies as measures of achievement (success rate) of the goals of a functional state, and to include behaviours other than movement (Fig. 1.1). Thus, if motivating currencies can be estimated given a state, fitness consequences can be evaluated with or without evidence of a specific behavioural response. This 'functional state approach' helps to frame the cross-disciplinary links between the motivating currency, proximate constraints and ultimate consequences of behaviour, and encourage the view that behavioural context is a signal, rather than a noise, variable that could potentially fill in knowledge and data gaps of individual-based approaches to population consequence (e.g. the PCAD framework, NRA 2005; in spatial

ecology, Schick et al 2008, Morales et al 2010; agent-based models, McLane et al 2011). It is not suggested that all components in the concept model must be explicit; rather, their omission can be stated and assumptions justified within a common conceptual framework. This conceptual model is used throughout this thesis to improve upon analysis frameworks and as a reference for discussing the assumptions and implications of results.

1.3 High-latitude sperm whale as a model species

Cetaceans face rapid changes in their marine habitat due to noise and chemical pollution, fisheries (stock exploitation, incidental catch), development of coastal areas, offshore energy and mineral explorations, and climate change (Reeves et al 2003). Interactions of these stressors, and in some populations also the impact of past large-scale hunting on present-day population size and genetic variability, may further impact the ability of populations to compensate for each change (Tyack 2009). Species with relatively slow life history traits and high parental investment, such as sperm whales, have evolved strategies to maximise individual survival and can be expected to rely largely upon phenotypic plasticity, including learning, to cope with repeated anthropogenic stressors (Sih 2013). Such species are therefore both challenging and important to understand the consequences of behavioural disturbance.

This thesis uses sperm whales in sub-arctic foraging grounds in Kaikoura, New Zealand and Northern Norway as relatively simple model systems where individuals spend most of their time alone and feeding (Teloni et al 2008, Oliveira et al 2013). Sperm whale males migrate to these productive high latitude waters and are nearly twice as large as females (Best 1979, Teloni et al 2008, Engelhaupt et al 2009). Male sperm whales do not take an active role in breeding in their first 20 years of life, and form loose bachelor groups that become smaller with age. In contrast, females are typically found in socially cohesive groups closer to the low-latitude breeding grounds. Females may conceive at the age of 9, giving birth every 4-6 years (gestation 15 months, lactation 2-13 years) (Whitehead 2003). The sexual segregation in distribution and movement patterns of sperm whales have been explained by different energetic requirements for growth, which may have evolved via sexual selection (Cranford 1999, Teloni et al 2008). Successful foraging is therefore likely to

particularly impact sperm whale male fitness in the high-latitude foraging grounds that are the focus of my thesis.

The sperm whale is a cosmopolitan species with a diverse diet of mainly meso- and bathypelagic cephalopods (Santos et al 1999, Santos et al 2002). Fish species are consumed occasionally, although high prevalence in stomach contents has been reported regionally (notably in Iceland, Martin and Clarke 1986, and in New Zealand, Gaskin and Cawthorn 1967). Analysis of stomach contents have revealed a diversity of ingested cephalopod species and life stages (e.g. Simon et al 2003), ranging from small chiroteuthids (~400g) to the giant squid (~400kg) (Clarke 1996). Ingested fish include medium-sized (0.3-3m) bottom-dwelling species from at least 55 different genera (Kawakami 1980). Diet composition varies by stranding site, sex and individual within a group (Evans and Hindell 2004, Marcoux et al 2007), and males appear to take more fish and generally larger prey items than females (Kawakami 1980). Sperm whales are thought to concentrate on the most abundant available prey (Evans and Hindell 2004), which regionally could result in a more monotonous or 'specialist' diet (e.g. jumbo squid in the Gulf of California, Davis et al 2007). Sperm whales are hence thought to be generalist feeders and may have somewhat wider niche breadths than other deep-diving marine mammals, including northern bottlenose whales and elephant seals (Whitehead et al 2003).

Sperm whales forage at depth (200-1000 m, Watwood et al 2006), facing trade-offs between time spent in prey patches and recovering oxygen stores at the sea surface (Boyd 1997). Costs of longer, deeper diving can increase non-linearly because of the cost of lost time to process lactate built up during anaerobic respiration (Houston and Carbone 1992, Kooyman and Ponganis 1998). Sperm whales spend more than 70% of their time in foraging dive cycles (dive + post-dive surfacing), and dives can last up to an hour (Watwood et al 2006). Interestingly, sperm whale surface recovery can be relatively unaffected by deeper and longer dives, both in terms of the post-dive surface duration (typically 4-10 min; Gordon 1987, Watwood et al 2006) and blow rates (e.g. Drouot et al 2004), indicating that they rely primarily on aerobic metabolism. Whichever exact mechanism limits sperm whale foraging dive schedule, individuals can be expected to be highly efficient at foraging at depth.

1.4 Bioacoustic measurements

1.4.1 Echolocation: an active sensory system in the deep-sea environment

Echolocation is an active sensory system where sound is used to probe the environment by processing of returning echoes. Echolocation as a means for finding prey evolved independently in odontocetes and in bats, possibly as a functional elaboration of active spatial sensing (Schnitzler et al 2003). Echolocation for prey generally consists of three broad temporal phases: 1) search phase utilising slow and fairly regular sampling rates, 2) approach phase that begins when a prey item is selected and rates are subsequently increased when moving towards the target, and 3) terminal 'buzz' phase during which a rapid burst of pulses are emitted at a short range (Griffin 1960). During search phase, both bats and captive odontocetes have been shown to emit a search pulse and wait for echo returns before emitting another pulse, thus avoiding ambiguity of echo returns, and decrease inter-emission interval on approach to targets (Madsen and Surlykke 2013). The difference between the two-way travel time (TWTT) of sound and the inter-click interval (ICI) has been termed "processing delay", which might reflect the time it takes for the animal to cognitively process echo information and prepare a response, i.e. delay in the vocal-motor feedback (Au 1993, Madsen and Surlykke 2013).

The detection and discrimination capabilities of marine and terrestrial echolocators are subject to similar sensory trade-offs despite the very different acoustic media (Madsen and Surlykke 2013, Wilson et al 2013). Sound travels much more efficiently in seawater that has much higher impedance (x5) and lower absorption than air. The higher impedance of seawater means that sound pressure levels are relatively easy to achieve, but smaller differences in the impedances between the medium and marine prey, such as squid, also reduce their relative target strength (Madsen and Surlykke 2013). The performance of any sonar system, whether biological or man-made, is limited by ambient noise in the environment, volume reverberation and clutter (unwanted echoes) from non-targets (Urick 1983). Shorter wavelengths (high frequencies) of echolocation signals absorb faster and therefore limit detection range, in air especially, but can improve spectral resolution and reduce backscatter. Longer signal duration of echolocation signals increases energy of the signal but inherently incurs a time cost. More directional emissions can achieve higher

received level from a target (echo level) while reducing clutter, but obviously diminish the size of perceived space. Conversely, higher source levels increase echo level, but also clutter and complexity of the auditory scene (Madsen and Surlykke 2013). These sensory trade-offs influence optimal echolocation systems and behavioural strategies. A degree of functional convergence might be expected across species detecting similar prey within similar acoustic habitats. For example, comparative studies of echolocating bats have found similarities in the intensity and duty cycle of echolocation signals between species operating under similar acoustic parameters, such as clutter-levels. These can be used to classify species into functional groups or 'guilds' which are independent of their phylogenetic relationships (Schnitzler et al 2003).

Odontocete species have also been shown to produce a range of distinct biosonar signals. However, their adaptive significance is debated (Baumann-Pickering 2013, Madsen and Surlykke 2013, Madsen et al 2013, Wilson et al 2013). Odontocete echolocation clicks can be broadly classified as: 1) short (< 150 µs) broadband clicks produced by most studied delphinids (Au 1993), 2) narrowband, high frequency (NBHF) clicks with longer durations (Madsen et al 2005a), 3) species-specific frequency-modulated (FM upsweep) search clicks of beaked whales (Johnson et al 2006, Baumann-Pickering 2013, Madsen et al 2013), and 4) lower-frequency and high power sperm whale clicks (e.g. Møhl et al 2003) (classification sensu Madsen et al 2005a). While NBHF clicks (Type 2) are produced by small shallow water species in *Phocoenidae* and *Cephalorhynchus*, the short and broadband clicks (Type 1) are more typical of delphinids inhabiting deeper waters. However, while *Kogia breviceps* carries the unique sound production apparatus and morphology of the Physeteroids, this small (~3m adult size) squid eating species also produces NBHF clicks, almost identical to those of the harbour porpoise (*Phocoena phocoena*) (Madsen et al 2005a). These adaptations appear inconsistent with the difference between the deep-diving hunting behaviour of *Kogia* and the predominantly shallower foraging of harbour porpoises. Madsen et al (2005a) suggest that both species groups are exploiting a low noise window at >100 kHz, while a potential anti-predatory explanation has also been proposed (‘acoustic crypsis’ hypothesis, Morisaka and Connor 2007, Morisaka 2012). Cryptic behaviour has also been suggested to explain apparent lack of communication signals at shallow dive depths (<170 m) of tagged

Blainville's beaked whales (Aguilar De Soto et al 2012). The centroid frequency of the FM clicks of different beaked whale species appear to correlate with the species' body size. This could reflect an anatomical constraint or be an adaptation for detection of smaller prey items (Baumann-Pickering 2013). In contrast, the sperm whale that is the largest of odontocetes, produces clicks both near surface and at depth (Watwood et al 2006), and their clicks can be detected at long ranges (up to 40 km, Barlow and Taylor 1997) with potential for a large communicative space (up to 60 km, Madsen et al 2002).

1.4.2 Monitoring sperm whale foraging behaviour from their clicks

Passive acoustic monitoring (PAM) of odontocete echolocation can, with careful measurement and interpretation, provide unique insights into their hunting behaviour and sensory strategies. Cetaceans' reliance on sound for feeding also raises concerns for anthropogenic noise in the ocean (e.g. Southall et al 2007, Tyack 2009), which has partly motivated the development of PAM systems.

Sperm whale vocalisations have been studied for over three decades to infer their underwater behaviour (Watkins and Schevill 1977, Best 1979, Gordon 1987); however, it is not until recent advances in onboard acoustic recording tags that also measure depth and acceleration (DTAG, Johnson et al 2009) that have provided convincing evidence that sperm whales use echolocation to detect prey both at long and short range. Three lines of evidence link sperm whale clicks to echolocation: 1) the hypertrophied nose of the sperm whale appears anatomically analogous to the melon (Cranford 1999) of other odontocetes whose echolocation abilities have been verified in captivity (Au 1993); 2) evolutionarily and ecologically similar beaked whales have been shown to echolocate on prey by onboard acoustic recording of return echoes prior to terminal echolocation phase (Johnson et al 2004), and finally 3) association of sperm whale buzzes with depth modulation and increased manoeuvring during bottom phases of dives (Miller et al 2004). Nevertheless, it is possible that prey are also detected visually using downwelling light or bioluminescent cues (Fristrup and Harbison 2002).

Sperm whales emit regularly (~1 s) spaced search clicks termed 'usual' or 'regular' clicks, interspersed with rapid series of buzz clicks (or 'creaks', Gordon 1987). The high source

levels (>230 dB re: 1 μPa), low absorption (low frequency), low repetition rates and high directionality of sperm whale regular clicks maximise sonar range for prey detection (Møhl et al 2000, Møhl et al 2003). With a 15 kHz centroid frequency (Møhl et al 2003), regular clicks can be reflected from targets 2.5 cm in radius, and even smaller targets can provide efficient backscatter if they contain air bubbles (Madsen et al 2002). Sonar equation modelling shows that these high-powered regular clicks have the potential to detect prey patches at ranges of more than 500 m (Møhl et al 2003). In contrast, higher repetition rate and lower amplitude of buzzes appear to be functionally analogous to short range terminal echolocation in other odontocetes (Au 1993, Johnson et al 2004) and bats (Madsen and Surklykke 2013). Faster clicking during terminal echolocation is thought to increase the temporal resolution of the auditory scene (Madsen and Surlykke 2013). Sperm whales are able to alter their acoustic output by at least 20 dB, and the amplitude of regular clicks is positively correlated with the frequency, as has been observed in small odontocetes (Madsen et al 2002 b). Less is known about variability of buzzes, and whether any such heterogeneity could relate to a particular prey type or approach such as detection and ambush distance (Chapter 5). Sperm whales also produce less directional and infrequent slow clicks (Oliveira et al 2013) and stereotyped click patterns termed 'codas' (sensu Watkins and Schevill 1977) that are thought to have a social function, such as maintaining social cohesion (Whitehead 2003).

As initially proposed by Norris and Harvey (1972), there is growing evidence that the initial sound pulse is produced in the monkey lips in the foremost part of the nasal complex, homologous with the phonic lips in delphinids (Cranford et al 1996, Møhl 2001, Madsen et al 2003, Zimmer et al 2005). The sound reflects back within the skull, giving rise to multiple pulses within a click when recorded off the acoustical axis. Most of the sound energy from the phonic lips is projected towards the back of the skull where it is reflected forward into the junk of the spermaceti organ (the 'bent-horn model' sensu Møhl 2001). The sound is transmitted through the junk into the seawater, producing the first 'p1' pulse that dominates on-axis recordings (Zimmer et al 2005). Sound transmission experiments with fresh sperm whale carcasses confirm that analogous train of pulses with decaying amplitude and fixed time interval can be generated by injecting single pulses to the distal sac (Møhl

2001). Thus the inter-pulse interval (IPI) is a function of the spermaceti organ, and using an allometric relationship can be used to estimate body length (Gordon 1991, Growcott et al 2011).

The ability to interpret regular clicks as searching for prey and buzzes as attempts to capture prey opens an avenue for measuring foraging effort for sperm whales. However, the reliable monitoring of buzzes using surface hydrophones is made difficult by their apparent variability at the source (Watwood et al 2006, Miller et al 2009) and low received levels in field recordings (Madsen et al 2002). Directionality of buzzes and changing geometry between the whale and recorder may result in further orientation-dependent effects. Using animal-attached acoustic tags, the detection probability for a buzz emitted by the tag individual can be assumed constant and close to one, when ambient noise levels are accounted for (Chapters 3-5). The disadvantages of acoustic tags are that they are expensive and relatively restricted in deployment duration and number of individuals sampled. They may also affect the behaviour of the animals to which they are attached (Chapter 3). Towed hydrophones can be simpler to use in many situations and can be used to sample both at population and individual level (Chapter 1).

Apart from using buzzes as an indicator of prey-capture attempts (Miller et al 2004), there have been several other attempts to measure sperm whale feeding success indirectly. Defecation rates have been shown to vary between regions with different oceanographic features (e.g. Whitehead 2008); however, defecation rates may not be comparable due to differential digestion of prey and the time lag between ingestion and defecation is not known. More indirectly, foraging theory predicts that patch residence time should increase at high quality patches (e.g. Pyke 1984, Houston and Carbone 1992). For sperm whales, patch residency could be measured in two-dimension as area-restricted search (sensu Kareiva and Odell 1987) or in three dimension as time spent in a dive, or a specific foraging phase within a dive (e.g. Miller et al 2004, Watwood et al 2006). The use of these indirect measures in parallel with buzz rates could provide a means to cross-reference their ability to predict feeding success and habitat preference. Indeed, prey encounter rates could provide a useful measure of small-scale habitat preference and alternative to presence-only techniques where absence data is not available.

1.5 Scope and organisation of thesis

This thesis aims to quantify behavioural responses to anthropogenic disturbance by using proxy currencies that can be more directly linked to individual fitness than use of generic behaviour time series. Three main themes are explored: proxies of foraging success and movement cost within behavioural modalities (all chapters), research effects of focal follows and tagging operations (Chapters 2, 3), and relating any behavioural effects to specific anthropogenic disturbance stimuli (Chapters 2, 4). Use of terminal echolocation buzzes as a proxy for foraging success is a continual theme throughout the thesis, with a specific focus for improving upon its use and interpretation in the final Chapter 5.

Chapter 2 “effects of whale-watching” provides a case study where behavioural responses are measured as a range of metrics related to the whales’ acoustic behaviour – time to first click, initial mean click interval, time to first buzz, buzz rate, and so on. While the multivariate approach was able to account for spatiotemporal variation in the response metrics and small changes in echolocation patters were detected, no significant effects of the whale-watching activity on foraging success could be inferred.

Chapter 3 “effects of tagging” improves upon the metric-by-metric treatment of acoustic behaviour by developing a novel hidden state model that used prior biological knowledge and multiple streams of DTAG data to estimate behavioural states that could be directly linked to functional behaviours at dive scale (searching at a prey-layer, resting, and recovery at surface). Both the second and first chapters quantify research effects in order to more accurately estimate any behavioural effects to the stimuli of interest.

Chapter 4 “effects of sonar” utilized the controlled exposure experiment (CEE) conducted by the Sonar, Sea Mammals, Safety (3S) collaboration in Norway to make a quantitative contrast between behavioural changes during experimental sonar, incidental sonar, and experimental controls. The time series of functional states estimated in Chapter 3 were used to test three hypotheses for behavioural change with a playback of mammal-eating predator sounds as a biologically relevant positive control: 1) change in functional time budget, 2) change in foraging success, given a state, and 3) change in locomotion cost, given a state. Convolving the effects on functional behaviours (time cost) from the two energetic

proxies allowed for a more accurate inference of the disturbance stimuli due to expectation of context-specificity of physical effects vs. perceptual biases.

Chapter 5 “linking buzzes to prey” examines sperm whale movement and sensory tactics during prey encounters more closely to better understand the observed variability in buzz rates (Chapters 2-4). Three hypotheses were tested: 1) a collection of movement tactics are employed to capture different prey types; 2) depth influences prey selection and therefore movement strategies; and 3) sensory tactics are related to movement strategies and prey types. Results from this chapter are used to recommend that depth is included in models for prey encounter rate, and that changes in prey encounter rate are best examined specific to context.

Tables and figures can be found at the end of each chapter following references (e.g. Fig. 1.1). Electronic appendices are cited using the Chapter number first, following alphabetical sub-section (e.g. 2A).

1. References

Aguilar de Soto N, Madsen PT, Tyack P, Arranz P, Marrero J, Fais A, Revelli E, Johnson M (2012) No shallow talk: Cryptic strategy in the vocal communication of Blainville’s beaked whales. Mar Mammal Sci 28:E75–E92

Au W (1993) The sonar of dolphins. Springer-Verlag, New York, US

Barlow J, Gisinger R (2006) Mitigating , monitoring and assessing the effects of anthropogenic sound on beaked whales. J Cetacean Res Manag 7:239–249

Barlow J, Taylor B (1997) Acoustic census of sperm whales in the eastern temperate North Pacific. J Acoust Soc Am 102:3213–3213

Baumann-Pickering S, McDonald M a, Simonis AE, Solsona Berga A, Merkens KPB, Oleson EM, Roch M a, Wiggins SM, Rankin S, Yack TM, Hildebrand JA (2013) Species-specific beaked whale echolocation signals. J Acoust Soc Am 134:2293–301

Beale CM (2007) The behavioral ecology of disturbance responses. Int J Comp Psychol 20:111–120

Bejder L, Samuels A, Whitehead H, Finn H, Allen S (2009) Impact assessment research: use and misuse of habituation, sensitisation and tolerance in describing wildlife responses to anthropogenic stimuli. Mar Ecol Prog Ser 395:177–185

Berger-Tal O, Polak T, Oron A, Lubin Y, Kotler BP, Saltz D (2011) Integrating animal behavior and conservation biology: a conceptual framework. Behav Ecol 22:236–239

Best PB (1979) Social Organization in Sperm Whales, Physeter macrocephalus. In: Behaviour of Marine Animals (eds. Winn HE, Olla BL). Plenum press, New York, US

Bindra D (1978) How adaptive behavior is produced: A perceptual-motivational alternative to response-reinforcement. Behav Brain Sci 1:41–91

Biuw M, McConnell B, Bradshaw CJA, Burton H, Fedak M (2003) Blubber and buoyancy: monitoring the body condition of free-ranging seals using simple dive characteristics. J Exp Biol 206:3405–3423

Boyd I (1997) The behavioural and physiological ecology of diving. Trends Ecol Evol 12:213-217

Brown JS, Kotler BP (2004) Hazardous duty pay and the foraging cost of predation. Ecol Lett 7:999–1014

Clarke M (1996) Cephalopods as prey. III. Cetaceans. Philos Trans R Soc Lond B Biol Sci 351:1053–1065

Cooke SJ, Hinch SG, Wikelski M, Andrews RD, Kuchel LJ, Wolcott TG, Butler PJ (2004) Biotelemetry: a mechanistic approach to ecology. Trends Ecol Evol 19:334–43

Cranford T (1999) The sperm whale's nose: sexual selection on a grand scale? Mar mammal Sci 15:1133–1157

Cranford T, Amundin M, Norris K (1996) Functional morphology and homology in the odontocete nasal complex: implications for sound generation. J Morphol 228:223–285

Curé C, Antunes R, Alves AC (2013) Responses of male sperm whales (Physeter macrocephalus) to killer whale sounds: implications for anti-predator strategies. Sci Rep 3:1579

Davis R, Jaquet N, Gendron D, Markaida U, Bazzino G, Gilly W (2007) Diving behavior of sperm whales in relation to behavior of a major prey species, the jumbo squid, in the Gulf of California, Mexico. Mar Ecol Prog Ser 333:291–302

Dingemanse NJ, Wolf M (2010) Recent models for adaptive personality differences: a review. Philos Trans R Soc Lond B Biol Sci 365:3947–58

Drouot V, Gannier A, Goold J (2004) Diving and feeding behaviour of sperm whales (Physeter macrocephalus) in the northwestern Mediterranean Sea. Aquat Mamm 30:419–426

Dunlop R a, Noad MJ, Cato DH, Kniest E, Miller PJO, Smith JN, Stokes MD (2013) Multivariate analysis of behavioural response experiments in humpback whales (Megaptera novaeangliae). J Exp Biol 216:759–70

Ellison WT, Southall BL, Clark CW, Frankel AS (2011) A new context-based approach to assess marine mammal behavioral responses to anthropogenic sounds. Conserv Biol 26:21–8

Engelhaupt D, Hoelzel a R, Nicholson C, Frantzis A, Mesnick S, Gero S, Whitehead H, Rendell L, Miller P, Stefanis R De, Cañadas A, Airoldi S, Mignucci-Giannoni A a (2009) Female philopatry in coastal basins and male dispersion across the North Atlantic in a highly mobile marine species, the sperm whale (Physeter macrocephalus). Mol Ecol 18:4193–205

Evans PGH, Hammond PS (2004) Monitoring cetaceans in European waters. Mamm Rev 34:131–156

Evans K, Hindell M (2004) The diet of sperm whales (Physeter macrocephalus) in southern Australian waters. ICES J Mar Sci 61:1313–1329

Frid A, Dill L (2002) Human-caused disturbance stimuli as a form of predation risk. Conserv Ecol 6: 11

Fristrup K, Harbison G (2002) How do sperm whales catch squids? Mar Mammal Sci 18:42–54

Gaskin D, Cawthorn M (1967) Diet and feeding habits of the sperm whale (Physeter catodon L.) in the Cook Strait region of New Zealand. New Zeal J Mar Fresh1:156-179

Getz WM, Saltz D (2008) A framework for generating and analyzing movement paths on ecological landscapes. Proc Natl Acad Sci USA 105:19066–71

Gill J, Norris K, Sutherland W (2001) Why behavioural responses may not reflect the population consequences of human disturbance. Biol Conserv 97:265–268

Gordon J (1987) The behaviour and ecology of sperm whales off Sri Lanka. PhD diss., University of Cambridge

Gordon JCD (1991) Evaluation of a method for determining the length of sperm whales (Physeter catodon) from their vocalizations. J Zool 224:301–314

Griffin D, Webster F, Michael C (1960) The echolocation of flying insects by bats. Anim Behav 8:141-154

Growcott A, Miller B, Sirguey P, Slooten E, Dawson S (2011) Measuring body length of male sperm whales from their clicks: The relationship between inter-pulse intervals and photogrammetrically measured lengths. J Acoust Soc Am 130:568–573

Hotchkin C, Parks S (2013) The Lombard effect and other noise-induced vocal modifications: insight from mammalian communication systems. Biol Rev Camb Philos Soc 88:809-824

Houston AI, Carbone C (1992) The optimal allocation of time during the diving cycle. Behav Ecol 3:255–265

Huggenberger S (2014) An acoustic valve within the nose of sperm whales Physeter macrocephalus. Mamm Rev 44:81–87

Johnson M, Aguilar de Soto N, Madsen PT (2009) Studying the behaviour and sensory ecology of marine mammals using acoustic recording tags: a review. Mar Ecol Prog Ser 395:55–73

Johnson M, Madsen PT, Zimmer WMX, Soto NA de, Tyack PL (2004) Beaked whales echolocate on prey. Proc R Soc B-Biological Sci 271 Suppl:S383–6

Johnson M, Madsen PT, Zimmer WMX, Soto NA de, Tyack PL (2006) Foraging Blainville's beaked whales (Mesoplodon densirostris) produce distinct click types matched to different phases of echolocation. J Exp Biol 209:5038–5050

Kareiva P, Odell G (1987) Swarms of predators exhibit "preytaxis" if individual predators use area-restricted search. Am Nat 130:233–270

Kawakami T (1980) A review of sperm whale food. Sci Reports Whales Res Inst: 32:199–218

Kooyman GL, Ponganis PJ (1998) The physiological basis of diving to depth: birds and mammals. Annu Rev Physiol 60:19–32

Lusseau D, Bejder L (2007) The long-term consequences of short-term responses to disturbance experiences from whalewatching impact assessment. Int J Comp Psychol 20:228–236

Madsen PT (2005) Marine mammals and noise: Problems with root mean square sound pressure levels for transients. J Acoust Soc Am 117:3952

Madsen PT, Carder D a., Au WWL, Nachtigall PE, Møhl B, Ridgway SH (2003) Sound production in neonate sperm whales (L). J Acoust Soc Am 113:2988-2991

Madsen P, Carder D, Bedholm K, Ridgway S (2005) Porpoise clicks from a sperm whale nose—Convergent evolution of 130 kHz pulses in toothed whale sonars? Bioacoustics 15:195-206

Madsen PT, Johnson M, Soto NA de, Zimmer WMX, Tyack P, Aguilar de Soto N (2005) Biosonar performance of foraging beaked whales (Mesoplodon densirostris). J Exp Biol 208:181–194

Madsen PT, Soto NA de, Arranz P, Johnson M (2013) Echolocation in Blainville's beaked whales (Mesoplodon densirostris). J Comp Physiol A Neuroethol Sens Neural Behav Physiol 199:451–69

Madsen PT, Surlykke A (2013) Functional convergence in bat and toothed whale biosonars. Physiology 28:276–83

Madsen PT, Wahlberg M, Møhl B (2002) Male sperm whale (Physeter macrocephalus) acoustics in a high-latitude habitat: implications for echolocation and communication. Behav Ecol Sociobiol 53:31–41

Marcoux M, Whitehead H, Rendell L (2007) Sperm whale feeding variation by location, year, social group and clan: evidence from stable isotopes. Mar Ecol Prog Ser 333:309–314

Martin A, Clarke M (1986) The diet of sperm whales (Physeter macrocephalus) captured between Iceland and Greenland. Mar Biol Assoc United Kingdom:779–790

McLane AJ, Semeniuk C, McDermid GJ, Marceau DJ (2011) The role of agent-based models in wildlife ecology and management. Ecol Modell 222:1544–1556

Miller PJO, Johnson MP, Madsen PT, Biassoni N, Quero M, Tyack PL (2009) Using at-sea experiments to study the effects of airguns on the foraging behavior of sperm whales in the Gulf of Mexico. Deep Sea Res Part I Oceanogr Res Pap 56:1168–1181

Miller PJO, Johnson MP, Tyack PL (2004) Sperm whale behaviour indicates the use of echolocation click buzzes "creaks" in prey capture. Proc R Soc London Ser B Biol Sci 271:2239–47

Møhl B (2001) Sound transmission in the nose of the sperm whale Physeter catodon . A post mortem study. J Comp Physiol A Sensory, Neural, Behav Physiol 187:335–340

Møhl B, Wahlberg M, Madsen PT, Heerfordt A, Lund A (2003) The monopulsed nature of sperm whale clicks. J Acoust Soc Am 114:1143

Møhl B, Wahlberg M, Madsen PT, Miller LA, Surlykke A (2000) Sperm whale clicks: directionality and source level revisited. J Acoust Soc Am 107:638–48

Moore MJ, Early G a (2004) Cumulative sperm whale bone damage and the bends. Science 306:2215

Morales JM, Moorcroft PR, Matthiopoulos J, Frair JL, Kie JG, Powell R a, Merrill EH, Haydon DT (2010) Building the bridge between animal movement and population dynamics. Philos Trans R Soc Lond B Biol Sci 365:2289–301

Morisaka T (2012) Evolution of Communication Sounds in Odontocetes: A Review. Int J Comp Psychol 25:1–20

Morisaka T, Connor RC (2007) Predation by killer whales (Orcinus orca) and the evolution of whistle loss and narrow-band high frequency clicks in odontocetes. J Evol Biol 20:1439–58

Nathan R, Getz W (2008) A movement ecology paradigm for unifying organismal movement research. Proc Natl Acad Sci 105:19052–19059

National Research Council (2005) Marine mammal populations and ocean noise: determining when noise causes biologically significant effects. US National Academy of Sciences, Washington DC, US

Norris K, Harvey GW (1972) A theory for the function of the spermaceti organ of the sperm whale (Physeter catodon L). NASA Spec Publ 262:397

Nowacek DP, Thorne LH, Johnston DW, Tyack PL (2007) Responses of cetaceans to anthropogenic noise. Mamm Rev 37:81–115

O'Connor C, Norris D, Crossin G, Cooke S (2014) Biological carryover effects: linking common concepts and mechanisms in ecology and evolution. Ecosphere 5:1–11

Oliveira CC, Wahlberg M, Johnson M, Miller PJO, Madsen PT (2013) The function of male sperm whale slow clicks in a high latitude habitat: communication, echolocation, or prey debilitation? J Acoust Soc Am 133:3135–44

Patterson TA, Thomas L, Wilcox C, Ovaskainen O, Matthiopoulos J (2008) State-space models of individual animal movement. Trends Ecol Evol 23:87–94

Pyke G (1984) Optimal foraging theory: a critical review. Annu Rev Ecol Syst 15:523–575

Reeves R, Smith B, Crespo E, Notarbartolo di Sciara G (2003) Dolphins, whales and porpoises: 2002-2010 conservation action plan for the world's cetaceans. IUCN, Gland, Switzerland and Cambridge, UK

Romero LM (2004) Physiological stress in ecology: lessons from biomedical research. Trends Ecol Evol 19:249–55

Santos M, Pierce G (2002) Additional notes on stomach contents of sperm whales Physeter macrocephalus stranded in the north-east Atlantic. J Mar Biol Assoc UK:501–507

Santos M, Pierce G, Boyle P (1999) Stomach contents of sperm whales Physeter macrocephalus stranded in the North Sea 1990-1996. Mar Ecol Prog Ser 183:281–294

Schick RS, Loarie SR, Colchero F, Best BD, Boustany A, Conde D a, Halpin PN, Joppa LN, McClellan CM, Clark JS (2008) Understanding movement data and movement processes: current and emerging directions. Ecol Lett 11:1338–50

Schnitzler H-U, Moss CF, Denzinger A (2003) From spatial orientation to food acquisition in echolocating bats. Trends Ecol Evol 18:386–394

Sih A (2013) Understanding variation in behavioural responses to human-induced rapid environmental change: a conceptual overview. Anim Behav 85:1077–1088

Simon MJ, Kristensen TK, Kinze CC, Tougaard S, Tendal OS (2003) Gonatus fabricii (Mollusca, Theuthida) as an important food source for sperm whales (Physeter macrocephalus) in the Northeast Atlantic. Sarsia North Atl Mar Sci 88:244–246

Southall BLB, Bowles AE, Ellison WT (2007) Marine mammal noise-exposure criteria: initial scientific recommendations. Bioacoustics 33

Stankowich T, Blumstein DT (2005) Fear in animals: a meta-analysis and review of risk assessment. Proc Biol Sci 272:2627–34

Taylor B, Gerrodette T (1993) The uses of statistical power in conservation biology: the vaquita and northern spotted owl. Conserv Biol 7:489–500

Teloni V, Mark JP, Patrick MJO, Peter MT, Johnson MP, Miller PJO, Madsen, Madsen PT (2008) Shallow food for deep divers: Dynamic foraging behavior of male sperm whales in a high latitude habitat. J Exp Mar Bio Ecol 354:119–131

Tyack PL (2009) Acoustic playback experiments to study behavioral responses of free-ranging marine animals to anthropogenic sound. Mar Ecol Prog Ser 395:187–200

Tyack PL, Zimmer WMX, Moretti D, Southall BL, Claridge DE, Durban JW, Clark CW, D'Amico A, DiMarzio N, Jarvis S, McCarthy E, Morrissey R, Ward J, Boyd IL (2011) Beaked whales respond to simulated and actual navy sonar. PLoS One 6:e17009

Urick RJ (1984) Ambient noise in the sea. Peninsula Publishing, Los Altos, California, US

Watkins WA, Schevill WA, Schevill WE (1977) Sperm whale codas. J Acoust Soc Am 62:1485–1490

Watwood SL, Miller PJ, Johnson M, Madsen PT, Tyack PL (2006) Deep-diving foraging behaviour of sperm whales (Physeter macrocephalus). J Anim Ecol 75:814–825

Whitehead H (1996) Variation in the feeding success of sperm whales: temporal scale, spatial scale and relationship to migrations. J Anim Ecol 65:429–438

Whitehead H (2003) Sperm whales: social evolution in the ocean. University of Chicago press, Chicago, US

Whitehead H, MacLeod CD, Rodhouse P (2003) Differences in niche breadth among some teuthivorous mesopelagic marine mammals. Mar Mammal Sci 19:400–406

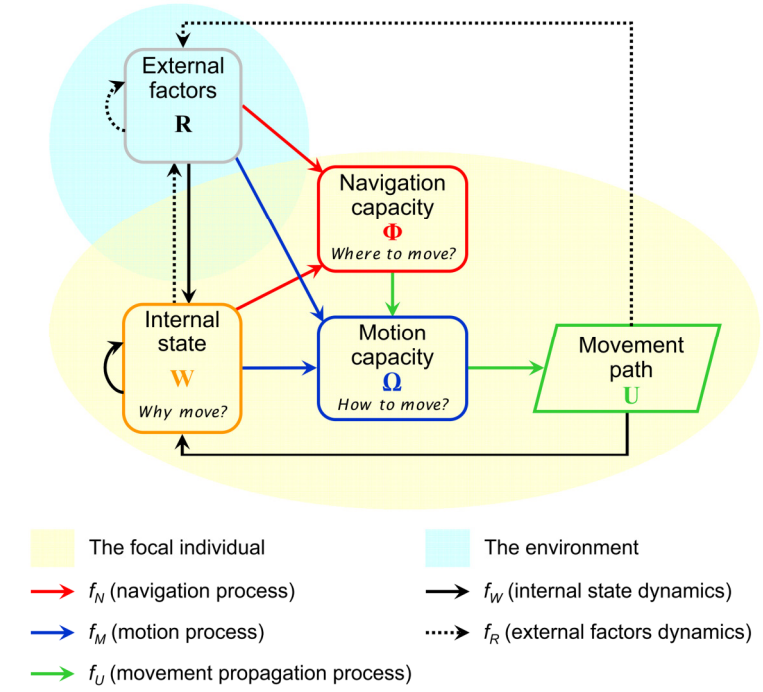
Wilson M, Wahlberg M, Surlykke A, Madsen PT (2013) Ultrasonic predator-prey interactions in water-convergent evolution with insects and bats in air? Front Physiol 4:137

Wilson RP, White CR, Quintana F, Halsey LG, Liebsch N, Martin GR, Butler PJ (2006) Moving towards acceleration for estimates of activity-specific metabolic rate in free-living animals: the case of the cormorant. J Anim Ecol 75:1081–90

Wright AJ, Deak T, Parsons ECM (2011) Size matters: management of stress responses and chronic stress in beaked whales and other marine mammals may require larger exclusion zones. Mar Pollut Bull 63:5–9

Zimmer WM, Tyack PL, Johnson MP, Madsen PT (2005) Three-dimensional beam pattern of regular sperm whale clicks confirms bent-horn hypothesis. J Acoust Soc Am 117:1473–1485

1. Figures



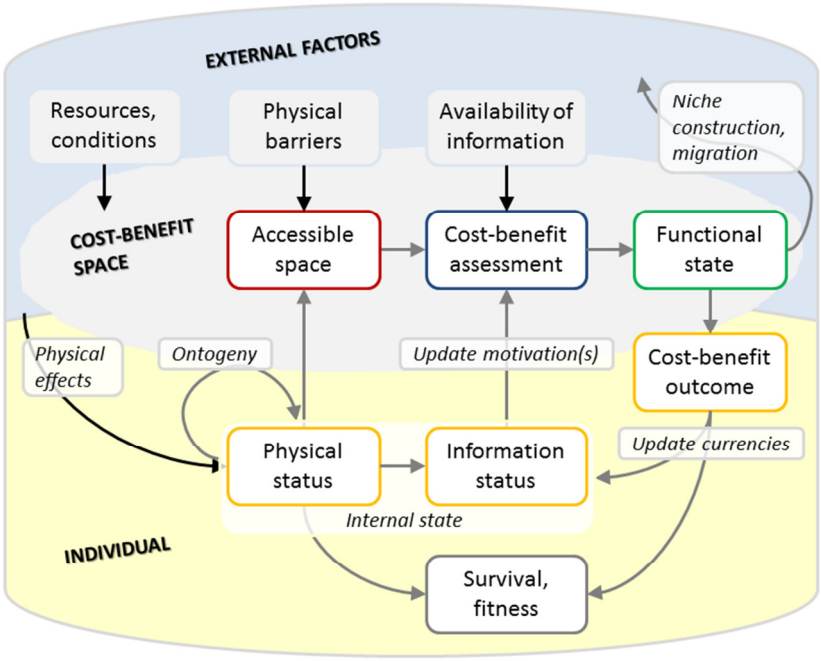


Figure 1.1 Movement ecology paradigm (top, Nathan et al 2008) and extended conceptual model for drivers of behavioural response (bottom)

The 'functional state approach' consists of the focal individual, its biotic and abiotic environment (external factors), and their interface (cost-benefit space). Cost-benefit space

is the outcome of behavioural options available to an individual, such as a trade-off between quantity and quality of offspring. Behavioural options are associated with a set of proximate goals or ultimate motivations (e.g. food, shelter). These options are limited physically and physiologically (accessible space) through individual history (ontogeny, e.g. sexual maturity), current physical status (reproductive state, body condition and homeostasis) and physical barriers. Cost-benefit assessment is the internal mechanism by which each individual selects a behavioural option. Information status encompasses cues, information and memory from both sensory and non-sensory inputs in the somatic nervous system. Functional state is the realized behavioural option and gives rise to a collection of behavioural traits that may influence the focal environment (niche construction). Currencies measure the actual cost-benefit outcome of the functional state and feedback to the internal state of the individual, thus with carry-over effects on subsequent behavioural options, functional states and fitness. While it is difficult to measure actual currency outcomes of behavioural choices for free-ranging animals, proxy indicators provide at least a relative indicator of costs and benefits.

2 EXAMINATION OF POSSIBLE EFFECTS OF WHALE-WATCHING ON SPERM WHALE ACOUSTIC BEHAVIOUR IN KAIKOURA CANYON

Abstract

Foraging is likely the most biologically significant activity for sperm whales that are the focus of whale-watching at Kaikoura, New Zealand. Foraging success is therefore a key parameter to monitor changes that might have a direct impact on the whales' fitness in the area. In this chapter, the acoustic foraging behaviour of whales was measured before, during and after viewing by different numbers and types of whale-watching vessels. No differences were found in acoustic parameters most closely related to prey encounter rates and feeding success that could be attributed to vessel interactions. There were slight changes to the initial search pattern of dives following whale-watching boat encounters: whales delayed their first click and descended for longer before stopping for their first silence, which was also extended. However, these effects appeared small relative to high individual variability and were within the range of behaviours predicted by their spatial and temporal environment.

2.1 Background

Whale-watching has grown rapidly to become a global industry (Hoyt 2001) and while frequently cited as "ecotourism" with many educational benefits, there is increasing concern about negative impacts on cetacean populations and lack of regulation (Corkeron 2004, Higham et al 2014). A widely held view of whale-watching as a non-consumptive alternative to whaling is disputed (Higham et al 2014), and is thought to underestimate potential non-lethal impacts (Lusseau and Bejder 2007) of an industry that has tremendous growth potential (Cisneros-Montemayor et al 2010). There is increasing evidence of both short-term behavioural effects, such as disruption of behavioural states (Lusseau 2003), as well as long-term impacts on local populations, such as habitat displacement (Lusseau and Bejder 2007). Off Kaikoura, sperm whales have been documented to alter both blow rates at surface and underwater echolocation behaviour, but their potential fitness impacts have been interpreted as likely minor (Gordon et al 1992, Richter et al 2006). Nevertheless, "transient" sperm whales that did not return to the area showed behavioural effects more consistently than "resident" whales, indicating more far-reaching effects such as habituation, and need for continual monitoring (Gordon et al 1992, Richter et al 2003, Richter et al 2006).

Male sperm whales migrate across ocean-basins to forage in productive high-latitude habitats such as Kaikoura. Sperm whales are highly sexually dimorphic (Cranford 1999) and presumably larger males are more reproductively successful. Their foraging success in feeding areas is therefore likely to be important for their future breeding success and survival. The submarine canyon in Kaikoura is one of the most productive benthic habitats known in the deep sea (De Leo et al 2010). Therefore, successful foraging is likely the most biologically significant activity for whales that are the focus of whale watching activities in Kaikoura.

While direct observation of sperm whale feeding is nearly impossible at typical foraging depths of 200-1200 metres (Watwood et al 2006), terminal echolocation buzzes or 'creaks' can be monitored remotely using hydrophones at the surface (Gordon 1987, Miller et al 2004). Sperm whale buzzes are rapid series of clicks where the initial inter-click-interval (ICI) of about 0.2 s decreases to about 0.02 s (up to 60 clicks per second; Goold and Jones 1995,

Madsen et al 2002a). The amplitude of buzzes is about 20 dB less than that of regular clicks (Madsen et al 2002a), and hence buzzes are more difficult to detect than regular clicks. Buzzes are variable in duration (5-30 s) and are usually followed by a short (2-10s) period of silence or regular clicks with longer ICI:s (Jaquet et al 2001, Madsen et al 2002a, Miller et al 2004) that could relate to recycling of the air in the sound generator, or handling captured prey. Although it is unlikely that all prey capture attempts are successful, buzz rates do match expected daily consumption rates of sperm whales (Whitehead 2003, Miller et al 2004). Unfortunately, there is, as yet, no means of remotely discriminating successful versus unsuccessful prey capture attempts in sperm whales.

This chapter investigates the acoustic behaviour of sperm whales while diving, with a focus on quantifying buzz production as an indicator of foraging. Sperm whale buzz rates have been found to vary greatly between individuals and on different measurement occasions (Miller et al 2009). This variability could be partly due to environmental factors (e.g. availability of prey) or behavioural mode (e.g. area-restricted search for prey vs. transiting). As buzzes are received at lower levels on surface hydrophones than regular clicks, and may be more directional, their detection probability will be lower than that of regular clicks, and hence more variable with changes in sound transmission between the source (whale) and receiver (hydrophone). Nevertheless, patterns on the more detectable regular click rates could also indicate encountering and approaching of profitable prey patches, and were therefore included in this chapter.

Individual whales were followed through several dives while making continuous acoustic recordings to measure changes in individual acoustic behaviour before, during and after they have been viewed by whale-watching vessels at the surface. In this way, factors other than whale-watching that could also have affected acoustic behaviour were controlled for. Specifically, four sources of variation were investigated in acoustic diving behaviour: 1) preceding surface behaviour, 2) spatiotemporal environment, 3) distance to research vessel during the preceding surfacing, and 4) presence of whale-watching boats during the preceding surfacing.

In the analyses of this chapter, a descriptive term ‘creak’ is adopted instead of the more functional term ‘buzz’ to highlight the fact that they were monitored remotely and both

their detection probability and properties such as duration could change at an unknown range and aspect to the whale. 'Buzz' is used to describe terminal echolocation as measured from DTAGs (Chapters 3-5).

2.2 Methods

2.2.1 Field methods

Field data collection aimed to acoustically track individual whales through complete dive cycles before, during and after being encountered by whale-watching (ww) vessels. Whales were first localised using a combination of towed hydrophones, a hand-held directional hydrophone, and radio communication with ww-operators in the area and land based observers. An omni-directional towed hydrophone was deployed once the dedicated research vessel ('Titi' 6 m monohull boat with 100 hp four-stroke engine) was in the vicinity of a surfaced or diving whale. The towed hydrophone was used to make continuous stereo recordings.

Two types of acoustic systems were used. First, hand-held directional hydrophones and headphone amplifiers were used to localise sperm whales. The second system was a towed hydrophone, and associated, signal conditioner, digitiser and computer used both as a means of finding and tracking sperm whales and for making continuous recordings throughout entire dives. The towed hydrophone was built by Ecologic UK and was based on their standard configuration. It consisted of a streamlined sensor unit (made up of 5 m of 35 mm oil filled polyurethane tube) towed on 100m of Kevlar strengthened cable. Within the sensor unit, and separated by 1 m, were two Magrec HP-03 spherical hydrophones with associated preamplifiers. These preamplifiers provide 29 dB of gain and had low cut filters set at 100 Hz to reduce low frequency flow noise. The streamer section also contained a Keller 10 bar 4-20 mA pressure sensor to provide information on hydrophone depth.

Signals from the hydrophone were amplified and conditioned on the vessel with a Magrec HP27ST stereo amplifier/filter unit and digitised by a RME Fireface 400 sound card at 96 or 192 kHz. The sound card was controlled by an Aeon Boxer fan-less 12v computer that ran PAMGUARD software to both carry out real time detection and to display sperm whale clicks, allowing real time tracking in the field, and to make continuous recordings to

hard drive. A USB GPS allowed PAMGUARD to collect location information and display tracks and detections on a real time map. The regulated 4-20 mA current from the pressure sensor induced a voltage over a 47-ohm resistor, which was measured by a Measurement Computing USB-1208 digital acquisition unit and converted to a depth reading, to be displayed and stored by PAMGUARD. The complete PAM system was 12 v powered.

At the surface, behaviour of the focal whale, including heading (first and last of a surfacing), timing of blows, start and end of surfacing (typically ending in a 'fluke-up' where the whale raised its flukes), and occurrence of any other identifiable behaviour were recorded using the Logger software provided by the International Fund for Animal Welfare. The presence of different types of commercial whale-watching vessels (boats, flights and helicopters) was recorded in logger if they were actively approaching and viewing the focal whale. Two visual observers were standing on a raised platform to allow a 360° view. One observer recorded timing of blows and fluke-up by pressing hotkey on a keyboard linked to the Logger computer, while the second observer was dedicated to taking photographs of whales' flukes for individual identification and communicating other visual data (Richter et al 2011).

Once the whale dove, it was followed usually by staying slightly behind its underwater location. The whale was tracked using its heading at fluke up as an initial guide, and the bearing to received clicks calculated and displayed in real time by PAMGUARD software. This program detected sperm whale clicks on digitised channels from each of the two hydrophones located 1 m apart in the hydrophone streamer, towed 100 m behind the vessel. Time of arrival differences for each click were processed to provide relative bearings to vocalising whales but with left-right ambiguity. These were displayed as a plot of bearing against time (e.g. Fig. 2.1). By monitoring this display the course of the research vessel could be adjusted to maintain effective tracking of the diving animal. To correct left-right ambiguities, when necessary, the vessel was stopped in order to take a bearing to the whale using the hand-held directional hydrophone.

2.2.2 Acoustic data processing

Visual data were used to identify long focal follows during which there were surfacing occasions with and without whale watching vessels. Continuous sound recordings during these follows were processed with PAMGUARD with a click detection module configured to automatically detect and localise clicks. The click detector doesn't classify the clicks to distinguish between those made by the focal whale clicks and clicks from other whales and other sources. Further, automatic detections often included false positives, especially as the thresholds were set low to detect as many quieter buzz clicks as possible. Clicks from the focal whale were therefore manually identified and marked using the Rainbow Click program.

Automatically detected clicks were stored as "click files" by PAMGUARD and these files were further analysed using Rainbow click. Rainbow Click has a bearing display similar to PAMGUARD (Fig. 2.1) and click train identification algorithms that identify clicks likely to be from the same individual based on their bearing and spectral characteristics. The operator's task was to review and edit these trains as necessary and link trains that were believed to come from the focal whale. The first clicks of dives (which are often characteristic in being loud and slow) were identified using the fluke up times noted in the visual dataset. Each surface start and end time was matched with the appropriate location in the respective sound files. Whales usually start regular clicking about 30-60 seconds after fluking and typically stop clicking some minutes before surfacing. They may emit a few slow clicks before and sometimes during surface periods (e.g. Fig 2.3).

To speed up analysis, files were not listened to in their entirety; instead, detected click trains were inspected visually and any identified gaps in regular clicking were listened to check for buzzes. Also any unusual/ slow-looking regular clicks were checked on, and aural monitoring was used to assess the total number of whales.

Clicks from focal whale were classified as being regular clicks (i.e., usual clicks), slow clicks, or creak (i.e., buzz) clicks. Clicks were marked as slow if they had an unusually long inter-click interval (ICI) (4-8s), were louder or sounded more metallic than previous/following regular clicks (Fig. 2.3). Clicks were marked as creak clicks when the ICI

was less than 100 ms and when the click train sounded continuous (i.e. the ear could not distinguish a click from another).

The automated click detector detected nearly 100% of regular clicks; however, its performance was much poorer for creak clicks. When gaps in clicking were carefully monitored by ear it was clear that a significant portion of creak clicks and even entire creak trains went undetected by the click detector. Thus, for consistency, the last and first regular clicks were marked before and after a creak, respectively (Fig 5.1, Fig 5.4) (hereafter termed 'creak interval'). If a creak was not heard, the pause was assumed not to have contained a creak, and was scored as a "silence".

Regular click trains and creaks produced by other whales were marked as individual click trains where possible. When these trains could not be linked with near certainty they were grouped or linked to each other with a "confidence" score of 0-100. This procedure ensured that the total number of animals heard at any given time was recorded while retaining as many click patterns as possible.

2.2.3 Summarising acoustic behaviour

A set of acoustic parameters was chosen to characterise the acoustic behaviour of the whale while diving (Table 2.1). Most of the parameters were directly comparable to those used in a previous whale-watching impact study in Kaikoura (Gordon et al 1992).

Whales could not be tracked successfully for all dives. For example, a dive became incomplete when the click train of the focal whale was so similar in bearing and amplitude to those of other whales that its track was lost. The summary parameters that could be extracted for a dive therefore depended to some extent upon the period over which the focal whale's acoustics output could be followed. Parameters that summarise the initial phase of clicking (time from fluke to first clicks *firstRCTime*, duration of first bout of clicks *firstRCTrainDur*, duration of first silence *firstPauseDur*, ICI:s of the initial bouts *first5ICI* and *last5ICI*) could be extracted in all cases where clicks could be detected until the second bout of clicks into the dive. The shortest analysed periods were discarded for parameters that required longer analysis duration. These were creaks that typically occurred much later in the dive (10-30 minutes), and parameters that were used to describe the search phase

(regular clicking) of the dive. Therefore, only dives that exceeded a set threshold in duration were used in the analysis of these parameters. After inspecting the respective parameter distributions, a 30-minute threshold was chosen for analysing the number of creaks (creak activity $CIFAct$, creak rate $CIFRate$ and proportion of time spent creaking $durCI$), and a 20-minute threshold for overall ICI (median ICI $medICIRC$, proportion of clicks within 0.1 seconds of median ICI $propMedICIRC$ and proportion of time spent silent $durPauseRC$).

2.2.4 Covariate data

A set of environmental variables and parameters measured in the field and from published data (e.g. bathymetry) were used to explain the acoustic behaviour in the subsequent dive (these are listed in Table 2.2). The year and month of observation were also included in the models.

2.2.5 Measuring change in acoustic behaviour

Exposed dives were defined as dives that were made immediately after a surfacing where ww-vessel(s) were actively viewing the focal whale. To compare with the previous study (Gordon 1992), pairs of consecutive exposed and non-exposed dives were first tested using Wilcoxon matched-pairs signed-rank tests (package MASS in r).

To capture some of the natural variability in the acoustic parameters, they were modelled as a function of surface covariates, year and month, as well as the number and type of ww-vessels (Table 2.2). Generalised additive mixed models (GAMM) (package mgcv in r; Wood 2004) were used to allow for more flexible non-linear responses to explanatory covariates, and explicit modelling of serial correlation within each focal follow. Smoothing functions were applied to each explanatory covariate to allow for more flexible responses. The complexity of the smoothing functions was restricted by setting their knots to maximum five for all covariates. The models were additive, meaning that the each of the explanatory covariates was assumed to contribute to the value of the acoustic response parameters independently from each other.

Exposure to ww-vessels at the sea surface may influence behaviour during the subsequent dive through a direct effect or less directly as a consequence of changes in

whale surface behaviour, e.g. by hindering full recovery of oxygen stores. If surface behaviour was not influenced by whale-watching, natural surface behaviour may still influence the subsequent dive behaviour (e.g. direct effects due to masking of prey echoes). Three models were therefore applied:

1) Null model fit only to non-exposed dives, including all covariates but whale-watching. This model tests how the year, time of year, the spatial environment and surface behaviour may be related to the acoustic parameters.

2) ww-model fit to all data, including all covariates except surface behaviour. This model tests whether the acoustic behaviour was different after whale-watching presence during the previous surfacing, given the spatial and temporal environment.

3) Full model was fit to all data, including all covariates in Table 2.2. This model tests if the acoustic behaviour is different after whale-watching presence during the previous surfacing, given surface behaviour, and the spatial and temporal environment.

Individual was included as an explanatory factor covariate in the whale-watching and full models, but there were not enough non-exposed data to include individual in the null model. A first-order autocorrelation structure (AR1) was estimated for each model to account for serial correlation within focal follow occasions. The autocorrelation structure assumed independence between followed whales, and that sequential dives were equivalently correlated.

Shrinkage smoothers were used as a means of automated model selection (Marra and Wood 2011). In other words, variables retained after shrinkage were judged to be important in capturing the variability in the response data.

2.3 Results

2.3.1 Extracted data

Clicks were extracted from 76 different dive cycles (surfacing + following dive) from 22 follow occasions with at least 11 individuals (confirmed photo-id). Of these, 46 were complete dives (whale followed acoustically from fluke up to surfacing) from 18 follow occasions of at least 10 individuals. Whale watching vessels were present for 36 of the

surfacings preceding the dives (Table 2.3, Fig 2.2). See Figures 2.3 and 2.4 for example extractions.

2.3.2 Acoustic behaviour before, during and after exposure to whale-watching

We first compared acoustic behaviour between dives for the same tracked individuals that were 1) not exposed to any whale-watching, 2) only viewed by whale-watching boats and 3) only viewed by aeroplanes or helicopters in the preceding surfacing.

There were no apparent differences in the means of acoustic parameters between non-exposed dives and dives exposed to whale-watching boats alone; all differences in means were small compared to their standard deviations, parameters related to creaks especially so (Table 2.4, Table 2.5, Appendix 2A). There were not sufficient data to investigate differences between non-exposed dives and dives exposed to ww-flights alone.

To control for variation in total acoustic activity, differences between pairs of subsequent dive cycles (surfacing + following dive) were tested with and without whale-watching exposure. All pairs were randomly sampled from different follow occasions. None of the parameters were found to follow a normal distribution, so non-parametric signed-rank tests were used. Three types of tests were carried out for each pair:

I. non-exposed dive vs. subsequent exposed dive
II. exposed dive vs. subsequent non-exposed dive
III. previous non-exposed dive vs. subsequent non-exposed dive, with an exposed dive in between

The three tests were carried out separately for boats and flights. To increase sample size, the presence/ absence of ww-boats and flights was considered independently of the presence/ absence of the other, i.e. surfacing with only ww-flights present would score as an absence of ww-boats. Even then, there were insufficient data to carry out the Type 3 test for the presence of ww-flights (Table 2.6).

The only significant test result (sig. level 0.05) was the duration of the first pause ($firstPauseDur$). The duration was increased by a median of 2.53 s for comparisons between dives with ww-boats absent in the first and present in the second preceding

surface period (Type 1 test, p=0.008, n=8 pairs). Similarly, *firstPauseDur* decreased by a median of 3.89 for comparisons between dives with ww-boats present in the first and absent in the second preceding surfacing (Type 2 test, p=0.054, n=11) (Table 2.6). This is consistent with the slightly higher mean *firstPauseDur* for all dives after they were viewed by ww-boats alone on the surface (11.20, sd=5.40), compared to dives that were not exposed to any ww-vessels (8.87, sd=4.65) (Table 2.4).

Although there was considerable variability in *firstPauseDur* between follows, the pattern appeared consistent within the follows. There was only one follow where the whale increased, rather than decreased its *firstPauseDur* in Type 2 test (exposed dive vs. subsequent non-exposed dive, Fig. 2.5). There appeared to be another small increase in *firstPauseDur* after two consecutive surfacings with boats present (Fig. 2.5), but paired testing did not show this to be significant (p=0.96, n=12).

Also the duration of the first bout of clicks (*firstRCTrainDur*) was increased between no-exposure and exposure (median difference +10.75 s, Type 1 test), and decreased after (-39.5 s, Type 2 test), but the p-values of the tests did not reach significant levels (p=0.31 and 0.24, respectively) (Fig. 2.6).

No significant differences could be detected in *firstPauseDur*, or any acoustic parameter, between previous non-exposed and subsequent non-exposed dives. However, the sample size was very small (3 pairs).

2.3.3 Model selection and modelled data

As described in the methods (Section 2.2.5), GAMMs were used to model the acoustic parameters as a function of environmental variables, surface behaviour and whale-watching while accounting for any serial correlation within focal follows (Table 2.2). For those initial click parameters that could be measured in all of the analysed data, all three models (null model, ww-model, and full model) were fitted to each parameter. Explanatory variables that captured a significant portion of the data were then automatically selected using shrinkage smoothers (Models 1-17 in Table 2.7).

For the parameters that described creaks and click intervals throughout the dive, there were insufficient data to fit the three models (Models 18-24 in Table 2.7). Because all

covariates could not be fitted within a single model, there was no full model from which to exclude candidate covariates. Unfortunately, step-wise model selection is not a reliable alternative in the mgcv library in R because the complexity of each response ('smooth') is estimated as part of model fitting and thus models are not nested within each other (Wood 2004). Instead, each parameter was modelled in turn with just one covariate, without an autocorrelation structure, and the importance of each covariate in the model was tested using Wald tests for parametric terms, and Bayesian 'p-values' for the smooths (anova.gam function in mgcv library in r). Covariates that were important at the 90% confidence level were included in a 'full' model that was then used in the automatic model selection by shrinkage (Marra and Wood 2011). The disadvantage of this approach is that any covariates that did not explain variability in the response data alone (at the 10% confidence level) were excluded from further analysis.

Appropriate model distributions and link functions were investigated for each parameter based on their information criteria (gcv score, Wood 2004), model convergence and resulting residual distributions. Residual distributions and model checking are given in Appendix 2B. Quasi-likelihood was used for proportions (Models 21, 23 and 24) and variables with over-dispersed distributions (Models 4-9 and 19, i.e. duration of first bout of clicks $firstRCTrainDur$, duration of first silence $firstPauseDur$ and creak activity *ClFAct*) (Table 2.7). ICI was best described as a Gaussian process; however, Gaussian models failed to converge for mean inter-click-interval (ICI) at the end of the first bout ($last5ICI$). The Tweedie distribution was used instead, which is implemented in mgcv as a mixed compound Poisson-Gamma distribution. Tweedie also performed well for time to first click, whose distribution has positive mass at zero. Time to first creak ($firstCTime$) was the only response variable that could be modelled as pure 'waiting time' Gamma distribution. Full model for the change in the initial ICI ($ICIChange$) did not converge with any of the tested distributions (Gaussian, Quasi or Tweedie).

2.3.4 Impact of the research vessel

To assess whether the research vessel itself could have influenced acoustic behaviour, minimum distance $minRange$ estimated on field to the whale during surfacing was used as a covariate. The covariate could not be included in the three full models due to small sample

size. Instead, two models were fitted for each acoustic parameter: both with *minRange* as a smooth non-linear covariate, the other also including presence of whale-watching boats as a second factor covariate. Furthermore, any remaining variability was checked for research vessel effects by modelling residuals of each model as a function of the minimum distance to the research vessel.

The *minRange* to the whale ranged from 25 to 1000 metres, with 64% of the data between 200-400 metres. When fitted alone, *minRange* did not explain sufficient data (at 90% significance level) in any of the acoustic parameters to be identified as an important covariate. *minRange* did not explain variability in any of the model residuals (at 95% confidence level) either. When the *minRange* was fitted together with presence of boats, *minRange* appeared to be important in the models for creak activity and proportion of time spent creaking (*durCI*). Both models were fitted with and without whale-watching boats that had little support in the two-covariate models (Wald tests p = 0.408 and 0.518, respectively), as well as with and without aspect and depth that were supported by model selection (Table 2.8 a).

The model for creak activity (*CIFAct*) appeared to be driven by the highest response value in the set, 20.1 creaks per hour that was measured after minimum range of 60 metres at the surface. Fitting without this point changed the fit so much that *minRange* was no longer supported when fitted with presence of boats (Wald test, n=29, p=0.377) and no longer converged in a model with aspect. This indicates that the model for *CIFAct* was over-fitting to the small sample size and that there may have not been real increase in *CIFAct* with proximity to the research vessel.

Presence of ww-boats was not supported in the models for the proportion of time spent in creak intervals (*durCI*, Wald tests p > 0.5), but both water depth and minimum range from the research vessel to the whale captured a significant amount of the data (Wald test, p < 0.001, n=30). The model with depth and *minRange* captured a large proportion of the data (adjusted R-square 63.62%) and it didn't appear to be driven by few data points or over-fit the data. Given mean depth, *durCI* was predicted to increase by a factor of 2.7 from *minRange* of 400 to 150 metres (Fig. 2.7).

2.3.5 Models describing the first bout of clicks

firstRCTime. Time to first click was 1 s or less in 28% of the modelled data, with a median of 7.25 s and maximum of 30.79 s. There was high between-individual variability with strong support for individual both in the ww-vessel and full models (Wald tests $p < 0.01$, Table 2.8 b). The full model explained 43.30% of the data, an increase of 6.18 units from the ww-vessel model which excluded surface covariates.

The aspect of seabed at fluking was an important predictor for *firstRCTime*, increasing with positive aspect (more west-facing slopes) in all three models ($p<0.04$, Wald test for Models 1-3, Table 2.8 a) (Figs. 2.8, 2.9). Water depth at fluke location was not retained in any of the three models, while bathymetric slope had weak support in the full model (Wald test, $p=0.289$, Table 2.8 a).

There was good support for heading change in the full model (Wald test, estimated df =0.89, n=59, $p=0.037$), but it failed to capture a significant amount of variability in the null model, probably because of much smaller sample size (n=33) and exclusion of the individual factor. A 160 degree change in heading was predicted to approximately halve the time to first click (Fig. 2.8).

Year and month were retained only in the null model, but this model did not include individual as an explanatory variable. Therefore, year and month were probably capturing variability that was better explained by the individual factor included in the ww-vessel and full models.

There was more support for the vessel covariate in the full model (Wald test, df=3, n=56, $p=0.023$) than in the ww-vessel model (n=59, $p=0.088$). Time to first clicks was significantly longer after encounters with boats or helicopters alone ($p=0.015$ and $p=0.030$, respectively Table 2.9); however, no such difference was detected after encounters with both boats and helicopters ($p=0.314$) (Table 2.9, Fig. 2.9, 2.10). Given mean values for all other covariates, no individual animals could be predicted to increase their time to first clicks after encounters with neither boats nor flights due to wide confidence intervals (95%) (Fig. 2.10).

firstRCTrainDur. The duration of the first bout of clicks was highly over-dispersed, with 67.11% of the data between 50 and 100 s, median of 80 seconds and maximum of 390 s (Fig.

2.11). The models accounted for this variation by estimating variance as a function of the mean (Table 2.7). The range of fitted values matched well with the range of the response values, and adjusted R-squares were over 50% for all three models (Table 2.7). There was strong support for an effect due to individual in both the ww-vessel and full models ($p<0.001$, Table 2.8 b).

Depth explained a significant amount of the data in the vessel and full model ($p<0.001$, Table 2.8 a), but was not retained in the null model, probably due to smaller sample size and not including individual as a factor. Duration of first bout of clicks increased in deeper fluking depths in both models, with a predicted difference of 43 seconds between 600 and 1200 metres, given the whale did not encounter vessels and mean values for all other covariates (Fig. 2.12 a).

Blows per minute (*blowRate*) prior to a dive captured a significant amount of the response data in both the null and full models ($p <0.001$, Wald tests Table 2.8 b). Higher blow rates were associated with an increase in the duration of first bout of clicks in both models, however, prediction intervals were too large to predict beyond blow rates of 4.5 (Fig. 2.12 b).

WW-boat presence was an important factor both in the ww-model and the full model ($p<0.004$, Table 2.8 b). Presence of boats alone was more important in the ww-model ($t=2.920$, $p=0.006$) than in the full model ($t=1.753$, $p=0.087$), while the presence of both boats and flights was more important in the full model ($t=2.991$, $p=0.005$) (Table 2.9). Neither model could show differences between no vessels and ww-flights, possibly due to small sample size. Not a single individual could be predicted to increase their time to first clicks after encounters with neither boats nor flights due to wide confidence intervals (95%) (Fig. 2.14).

firstPauseDur. The null model for duration of first silence did not retain any covariates and hence explained no data in the response variable. When individual was included in the vessel and full models, only fluking depth was retained with some support for the covariate ($p=0.07$, Wald tests Table 2.8 a), but the adjusted R square of both models remained low (7.40 and 8.28%, Table 2.7). There was little evidence that the presence of vessels explained significant amount data in the model (Wald test in Model 9, $n=56$, $p=0.368$) (Table 2.8 b).

first5ICI. Models for the initial ICI had the largest adjusted R-squares in the set (61-76%). The mean initial ICI appeared normally distributed, with a mean of 1.12 s, minimum of 0.73 and maximum of 1.65 s.

Water depth at fluke location was the most important covariate for the initial clicks in all three models ($p<0.001$, Table 2.8 a). A 0.12 second increase was predicted for every 100 metre increase in fluking depth (Figs. 2.15, 2.16a). Also slope was retained in the vessel and full model with good evidence that it explained the response data (Wald test $p<0.03$). A much smaller effect on initial ICI was revealed (compared to that of water depth) with about 0.1-0.2 second increase between flat and steep sea bottom at fluking (Fig. 2.16 b, d). In the null model, sea bottom aspect was retained instead of slope (Table 2.8 a).

There was strong support for month and individual in the ww-vessel and full models (Wald tests $p < 0.001$, Table 2.8 a). Initial ICI was predicted to be slightly longer in winter (May/August) than in spring (Sept-Nov); however, such an increase was not obvious in the raw data (Fig 2.15 c). The relationship was estimated as being linear due to lack of data for summer months of December-January, and only a few data points for February. These predictions should therefore be interpreted with caution. There was no evidence for vessel effects ($p>0.3$, Table 2.8 b, Table 2.9) (Fig. 2.15).

last5ICI. The full model explained much more of the ICI at the end of the first bout of clicks than the vessel and the null model (adjusted R square 57.32, 18.15 and 16.91%, respectively, Table 2.7). This is probably due to the model attempting to over-fit in a small data set where there was little actual signal. Indeed, the predicted relationships for fluking depth, heading change and surface duration appeared implausibly complicated in the full model (Fig. 2.17 a.), with the greatest changes predicted at the edges of the data. These predictions are therefore unreliable. Similarly the impact of vessel presence cannot be interpreted in the full model. There was little support for vessel presence in the ww-vessel model ($p=0.165$, Table 2.8 b).

Aspect of the seafloor was retained in all three models (Table 2.8 a). In all three models, *last5IC* was predicted to decrease with positive aspect (e.g. Fig. 2.17 b); however, the prediction intervals may be artificially narrow in the full model due to the over-fitting of data by the other covariates.

ICIChange. The full model for change in ICI did not converge with any explored distribution or link function. In the null and vessel model, there was strong support for water depth at fluking, but no other covariate ($p < 0.001$, Wald tests in Table 2.8 a). Change in ICI was predicted positive in shallower, and negative in deeper waters (Fig. 2.18).

2.3.6 Models describing creaks

firstCTime. The sample size used to fit the model for time to first creak was small (n=25), and the model explained only 13.68 % of the data (R-sq, Table 2.7). Sea bottom aspect at fluking position and number of encountered boats were retained in the model with weak support for aspect (p=0.208) and good support for number of boats (p=0.030) (Wald tests, Model 18 in Table 2.8). *firstCTime* appeared to decrease with number of boats. However, confidence intervals were too large and the model explained too little of the variation for reliable prediction.

CIFAct. Only sea bottom aspect was retained in the model for creak activity, explaining 18.46% of the variation in the data (n=30, Model 19, Tables 2.1, 2.2 a). The model fitted a considerably smaller and narrower range of values than the observed range of values (5.53-11.51 and 0-20.9 creaks per hour, respectively). According to the model, *CIFAct* (-h) increased with positive sea bottom aspect (Fig. 2.19).

CIFRate. Only sea bottom aspect explained enough variability in creak activity to be included in the full model, but it was not retained further by shrinkage. No covariates were therefore found that would have explained *CIFRate*.

durCI. Depth and individual explained 45.04% of the proportion of time spent creaking and fitted values matched well with the observed proportions 0-0.62 (Table 2.7). Only depth appeared to capture significant amount of *durCI* (p=0.043, Table 2.8a). For prediction, individual was retained in the model. Proportion of time spent creaking appeared much higher in deeper waters, but confidence intervals were large and the fit appeared to be driven by few unusually high proportions at the deeper range of the data (Fig. 2.20).

2.3.7 Models describing search click intervals

medICIRC. The distribution of median ICI appeared normal with a mean of 1.04, minimum of 0.61 and maximum of 1.53 s (n=38). The fitted values of the model matched almost exactly the observed range, explaining 85.57% (R-squared, Table 2.7) of the data. This is probably due to the ability of individual ($p<0.001$) to capture the variability.

Month, heading change and vessel presence were also included in the model with good support for month ($p=0.024$) and weaker support for presence of vessels ($p=0.089$) and heading change ($p=0.119$) (Table 2.8 a,b).

medICIRC was predicted to increase by 0.1 seconds from April to October, given mean values for all other covariates and no vessels present. The relationship was predicted linear as there were no data for December-January. Presence of vessels was predicted to increase median ICI by another 0.12 seconds, averaging across individuals and vessel factors (Fig. 2.21).

propMedICIRC. 51% of the proportion of clicks in the median ICI were less than 0.2, with a maximum of 0.6 (n=41). The model fit well to the data, with the range of fitted values matching the observed range of 0-0.6.

Fluking depth, sea bottom aspect, surface duration, blow rate and individual were retained in the model, explaining 45.15% of the data (R-squared, Table 2.7). Depth, aspect and individual were the most significant explanatory variables ($p<0.001$, Table 2.8 b), but there was also good support for surface duration and blow rate ($p=0.006$ and $p=0.018$) (Wald tests, Table 2.8 a).

Nearly linear relationships were estimated for each of the covariate, with *propMedICIRC* increasing with deeper fluking depths, longer surface durations and higher blow rates; *propMedICIRC* was predicted to decrease with positive sea bottom aspect (Fig. 2.23 b).

durPauseRC. The only covariate retained in the model for the proportion of time spent silent was sea bottom slope at fluking, with strong evidence that it explained variability in the data ($p<0.001$, Table 2.8 a). *durPauseRC* was predicted to increase when fluking at steeper slopes (Figure 6.22).

2.4 Discussion

This chapter analysed acoustic behaviour in order to explore changes in subsurface behaviour of whales that were likely to be of biological significance, foraging activity in particular. The challenge was to separate potential whale-watching effects from natural variability that is related to the measurement occasion due to individual, environment, location, season and year for example. First, pair-wise comparisons were made between dives for the same whale before and after they encountered whale-watching boats at the surface. These might be considered simple natural experiments, where mean differences between subsequent exposures could be effectively tested using non-parametric statistical tests. However, the requirement for sequential encounters of the same individual both with and without whale-watching limited the sample size obtained for this contrast. To allow a greater proportion of the collected data to be included in the analyses, an additional multivariate analysis was conducted where each of the acoustic parameters was modelled as a function of whale-watching and environmental variables, as well as individual. The disadvantage of this approach was that not all acoustic parameters fitted well to statistical distributions despite detailed examination of patterns in the data. The strengths and weakness of the two approaches were therefore somewhat complimentary.

2.4.1 Caveats of the analytical approach

The ability of the statistical analyses to detect change appeared somewhat limited by high variability between individuals and follow occasions and the relatively small sample size. Acoustic parameters related to number of creaks were particularly variable, and because they occurred later in dives, could only be analysed for a small portion of the data (25-30 dives, Table 2.7). Full models describing creaks and search click intervals were only fitted with covariates that could predict the response data in the absence of other covariates (Models 18-24, Table 2.7). In other words, not all explanatory covariates could be included in the full model, and therefore some features of the environment might have been omitted that could have explained smaller (but significant) amount of the response data, once other variability was accounted for. This is a disadvantage of the flexible model fitting approach for which step-wise selection of covariates is not appropriate (Wood 2004, Marra et al 2011). For example, only seafloor aspect was retained in the model for creak

activity (Model 19, Table 2.8). Thus, bathymetric aspect could be concluded to have captured a significant amount of the variability in creak activity, but not that it was the only one of the covariates that would have explained the data. It is possible that, for example, season would have captured a small amount of variability in creak activity once in a model with aspect. In contrast, for acoustic parameters describing the initial bout of clicks it can be assumed that the retained covariates were the only covariates capturing a significant amount of the data (Models 1-17, Table 2.8). Therefore, there was more confidence in a negative 'no whale-watching effect' result for the initial bout of clicks (Models 1-17) than for parameters describing creaks or search click intervals (Models 18-24). There was little difference between the two sets of models for any positive result.

Not all model predictions were considered reliable. Models for the ICI at the end of the first bout of clicks appeared to be over-flexible with respect to the small sample size (Models 13-15 in Table 2.7). This is a likely consequence of an over-flexible model fixing individual levels so that any residual variability can be fitted with the non-linear explanatory covariates. The predictions of the full model were implausibly complicated explaining nearly 60% of the data. However, a more positive bathymetric aspect was predicted to decrease the ICI in all three models. Given that the three models were fit to slightly different data sets, this added confidence to a conclusion that bathymetric aspect was an important explanatory variable.

Models for change in ICI during the initial bout did not converge well with any distribution, and although the null and ww-watching models (Models 16 and 17 in Table 2.7) could be fitted with a Gaussian distribution, the full model did not converge. This indicates that the Gaussian distribution may not describe the distribution of ICI Changes very well, and the model should therefore be interpreted with caution.

Any differences found between different seasons should also be interpreted with caution. There were no data for December-January, and as a result the models predicted nearly linear relationships between months (2-11) and the response variables, which seems unrealistic.

Due to the small sample size, whale-watching effects could only be investigated on the dive immediately following the whale-watching encounter. The models were not informed

about exposure to whale watching vessels on earlier dives. Similarly, the pair-wise tests only compared dives before and after encounters.

The detection of buzzes can be uncertain when monitoring sperm whales remotely using a towed hydrophone. When creaks were monitored aurally, it was apparent that some portions of creak trains were inaudible. In some cases it seemed that the beginning and end of the creak train was audible, but not creak clicks in between. The apparent silences are likely due to the whale changing the direction of the sound beam away from the hydrophone. Creak trains were typically preceded by faster regular clicks, but such faster regular clicks also lead to silences. This indicates that entire buzzes may have gone undetected. It is possible that features of the regular clicks could be used as an indicator of prey encounter rate, such as mean and variance of click rate.

2.4.2 Impact of the research vessel

The research vessel 'Titi' gave priority to whale-watching boats to view whales, but approached the whale from behind and attempted to remain at ranges > 300 m to collect surface and photo-identification data. Priority was determined based upon the New Zealand Marine Mammal Protection Regulations (1992), which limits the number of whale-watching vessels (whether commercial or recreational) to three within 300 metres of the whale. Unlike whale-watching boats, the research vessel attempted to track individual whales at very low speed during diving, and this could potentially be disturbing to the whales. There was no means to collect data of the distance between Titi and the diving whale, but there was great variability in the estimated distance to the whale at surface: minimum distance ranged from 25 to 1000 metres in the analysed data. Ranges less than 100 metres were measured on a few occasions (4 surfacings in the analysed data) when the whale surfaced next to the research vessel or swam towards the stationary vessel.

The only relationship between the acoustic parameters and the minimum distance between Titi and the whale during the previous surfacing was a slight increase in the proportion of time spent in creak intervals with closer proximity to the whale at surface. This is likely due to variability in buzz detection, rather than a real increase in the terminal echolocation of the whale. Observations of the whales' heading at fluke up and the fluking

location were used to determine the initial course of the research vessel when tracking a diving whale. These estimates were likely to be less reliable at greater observation distance. The whale's relative location was monitored with the towed and directional hydrophones, but it is possible that when the initial heading estimate was poor, more time was spent further away from the diving whale. Coupled with the directionality of buzzes, this could have degraded their probability of detection.

Overall, these results indicate that there were no obvious changes in the acoustic behaviour related to the research vessel's proximity to the whale at surface. However, minimum distance could not be fitted within the full models and therefore smaller effects cannot be ruled out. For the analysis of whale-watching impacts therefore, it is assumed that the impact of the research vessel was small and relatively constant effect during all encounters.

2.4.3 Acoustic behaviour in the environment

Most of the analysed data was collected within the Kaikoura Canyon and the northern part of Conway Trough. The Kaikoura Canyon is 60 km long and u-shaped in profile, joining Conway Trough at the head of the canyon only some 500 metres from shore (Lewis and Barnes 1999). The proximity of the canyon to land and the substantial sediment input may contribute to the Kaikoura Canyon being one of the most productive benthic habitats described so far in the deep sea (De Leo et al 2010). The canyon sediments are mostly mud, with the major gravel-sand-silt turbidity currents originating in the head of the canyon near Kahutara and Kowhai rivers (Lewis and Barnes 1999). Kaikoura Canyon also benefits from the subtropical convergence zone and nearly year-round upwelling of nutrient-rich waters. In winter, coastal upwelling of warm water is caused by mixing of river inputs and subtropical waters intruding into the canyon from southward flowing East Cape current (Houtman 1965). In summer, upwelling of cold water occurs when the north-flowing Southland Current converges with the more saline subtropical East Cape current against the continental shelf (Garner 1961, Heath 1971).

The oceanographic features of the Kaikoura Canyon and the surrounding waters parallel those used to describe sperm whale foraging habitat around the world's continental slopes

and ridges: high bottom relief (e.g. Jaquet and Whitehead 1996, Hooker et al 1999, Pirotta et al 2011), coastal upwelling of cold nutrient rich waters (Rendell et al 2004), thermal fronts (Griffin 1999) and high primary productivity (Jaquet and Whitehead 1996). These features are thought to result in concentrations of sperm whales' prey – mainly meso- and bathypelagic cephalopods, fish being more important regionally (Clarke 1996, Whitehead 2003).

Bathymetric features - depth, slope and aspect - were consistently indicated as being important in shaping the acoustic foraging behaviour of sperm whales in Kaikoura. Bathymetric features were retained in all full models related to the first bout of clicks, creak activity, proportion of time spent in creak intervals, proportion of clicks in median inter-click-interval (ICI) and proportion of time spent silent (Table 2.7, Table 2.8 a).

The echolocation behaviour in the beginning of the dive was related to the depth of the sea bottom. The mean ICI of the first five clicks was shown to increase by 0.12 s with every 100 m increase in water depth at the fluking location. Odontocetes click soon after they receive the echo of their target or the most distant large reflector (Au 1993). A change in water depth of 100 m would increase the travel distance for a click's echo by 200 m. If speed of sound was 1520 m/s, the travel time would increase by 0.13 s per 100m, very similar to the observed 0.12 s. Similarly, the initial ICI was close to the two-way travel time of sound to the bottom or slightly above depending on the individual and season. As slope steepness increased, there was a small increase in ICI. This could be explained if greater bottom depths were within the beam of ensonification than the depicted depth at that location (180x180m grid cell bathymetric resolution). Mean initial ICI was a median of 0.03 seconds shorter than expected by the two-way travel time, indicating a delay in the vocal-motor feedback (Au 1993, Madsen and Surlykke 2013). These results suggest that in the beginning of the dive, the whales waited for the returning echoes from near the seafloor before emitting another click. Jaquet et al (2001) had also found a correlation between fluking depth and the ICI of the first 10 clicks in Kaikoura, suggesting that the first clicks function to detect the sea bottom.

Similarly, the first pause was found to be slightly delayed in deeper waters. It has been suggested that pauses occur when air is recycled for sound production (e.g. Madsen et al

2002 b). The duration of the first bout, before the first pause, was predicted by the model to increase by about 7 seconds (or by 7-14 metres assuming a descent rate of 1-2m/s; Gordon 1987), for every 100 metre increase in sea bottom depth, given mean values for all other covariates (Fig 2.12 a). This result appears somewhat contradictory with compression of air at depth, which has been postulated to limit the number of clicks produced before recycling the air.

Clicks were more regular (higher proportion of clicks near median ICI) after longer surface durations, higher blow rates and in dives that started in deeper waters. Together with the delayed onset of the first silence, these results indicate that the whales undertook a different foraging strategy, possibly attempting to localise a deeper prey layer in deeper waters. The importance of longer surface duration and higher blow rate suggest that these dives are associated with an increased need to recover oxygen stores. Little evidence has been found for an increased need for recovery after longer or deeper dives (Watkins et al 2002, Drouot et al 2004, Watwood et al 2006, Davis et al 2007). This could be due to little correlation between dive duration and energy consumption, or that the studied whales did not routinely reach their aerobic dive limits. However, Jaquet (2001) found a weak but statistically significant correlation between dive duration and both the preceding and following surface duration in Kaikoura. In the same study, the correlation matched with a seasonal pattern: whales stayed longer at surface (+0.5 min) and made longer dives (+5.3 min) in the summer than winter. This coincided with whales congregating within the canyon, in relatively deep waters (>1000 m) during summers in 1990-1994. Similary, Richter et al (2003) documented longer surface durations during summer, but unlike Jaquet et al (2001), preference for the deepest part of the canyon in autumn and winter in 1994-2001. Unfortunately, there was little data available for this thesis chapter to detect seasonal changes in acoustic behaviour.

There was some indication that the initial ICI was longer in February-May and shorter in September-October (Fig. 2.16 c) than expected based on depth and slope alone. Similarly to Jaquet (2001) and Richter et al (2003), it is plausible that across the study period here there were seasonal changes in distribution and selection of prey that drive the observed changes in foraging strategy. Sperm whales are found in Kaikoura year-round and are likely to exploit

a range of available prey species. Sperm whales have been described as a generalist feeder (Whitehead 2003). For example Clarke (1996) found that sperm whales take advantage of aggregations of terminally spawning cephalopods. In Kaikoura, there are also reports of occasional surface feeding by the whale-watching operators (pers. comm. with Wings over Whales operators).

Regular clicks are typically faster before terminal echolocation (a buzz) (e.g. Jaquet et al 2001). More irregular usual clicking may therefore co-vary with buzz rates. In the present study, slopes facing west (positive aspect) were associated with both irregular clicking and higher creak activity (Figs 2.19, 2.23 a, 2.24). Similarly, the mean ICI at the end of the first bout was shorter when fluking over a sea bottom with a west facing slopes. These results indicate that the whales were detecting more prey at the coast-facing slope of Conway trough and south-west facing slope of the Kaikoura Canyon. This was likely related to the local water flow conditions at the shelf slope rather than the orientation of the slope itself. The dominant current in the study area is the Southland current flowing northwards through the Conway Trough, branching offshore south of the Kaikoura Peninsula (Heath 1971). West and south-west facing slopes therefore orientate nearly parallel to the current. The south-west facing slope of the Kaikoura coincides with the major route of sand-gravel-silt turbidity current flowing offshore from the head of the canyon (Lewis and Barnes 1999) and the summer congregations of sperm whales described by Jaquet et al (2000). Also Pirotta et al (2011) found aspect to capture more variability in the sperm whale distribution data than slope alone, in contrast to other studies where steepness of the slope was deemed important (e.g. Praca and Gannier 2007). These apparently contradictory results may be explained by differential importance of flow conditions, as well as the limited range of each study area that may encompass only a few combinations of slope and current.

Slope explained 34.7% of the proportion of time spent silent, with up to a two-fold increase in silent time between flat and steep sea bottom (Table 2.7, Fig. 2.22). This could be due to the whale spending more time listening passively to conspecifics or prey, especially if steep bottom relief altered sound paths to the whale.

A much smaller number of creaks was detected per dive (mean=4.17, sd=3.86, n=46 complete dives) than documented in the literature; for example Drouot et al (2004)

detected an average 24.8 creaks per dive. However, comparisons of creak rates across studies should be interpreted with caution because different studies might define creaks differently. Here, a creak was defined to start and end with a 100 ms threshold on ICI, but considered creaks joining with other creaks with few slower creak clicks as one.

Time to first click was 8.28 seconds on average (SD 7.27), considerably shorter and more variable than previously reported 31.9 and 25 s Kaikoura (Gordon et al 1992 and Jaquet et al 2001, respectively). The first click was delayed by 1-2 seconds with more positive aspect. This may be a reflection of the different foraging tactics the whales employ along different parts of the canyons' slopes. The model predicted a decreased time to first click with increased heading changes at surface, possibly related to an increased need to update information on bathymetry. This is supported by our field observations that whales tended to follow bathymetric contours up until a point in the canyon where they would turn back to follow the same contour lines back.

2.4.4 Acoustic behaviour and whale-watching

Duration of the first silence was the only acoustic parameter that showed significant and consistent differences with whale watching vessel presence in the non-parametric paired comparisons. The duration increased by a median of 2.53 seconds for dives made after being viewed by ww-boats, and decreased by a similar amount of 3.89 after a surfacing with no ww-boats. In the model for duration of first silence, however, all covariates failed to explain a significant amount of the data; bar weak support for depth ($p=0.07$, Table 2.8 a). The poor model fit could be explained by an inappropriate choice of distribution for the response; however, none of the tested distributions could find a signal in the data and the residual distributions appeared good for the quasi model (Appendix 2B). Another possibility is that the individual covariate captured enough variability in the response for any other explanatory covariate to become unimportant in the model. Duration of first silence is also the parameter that showed the strongest association with vessels in Gordon et al (1992). These results may indicate that some individuals remain silent for longer during this first pause when ww-vessels are present because they are listening passively to the vessels above them, perhaps assessing the directions in which they are moving away.

Vessel presence was a significant predictor of time to first clicks, along with individual, aspect of the sea bottom and the whale's heading change at surface. Presence of whale-watching boats and flights appeared to increase time to first click, and more significantly so after accounting for the heading change. Similarly, Richter et al (2003) found a 50% increase in time to first clicks, but only for transient animals; residents decreased their time to first click in the presence of whale-watching boats. It is possible that some whales delayed clicking until sufficiently far from any engine noise to start echolocation at ambient noise levels, or that they used this time to listen to the vessels moving off at surface.

Interestingly, Richter et al (2003, 2006) also found that “resident” whales showed greater and more frequent heading changes in the presence of whale-watching platforms. Greater heading changes decreased time to first click in the data, an opposite effect to the presence of whale-watching boats alone. If this reflected a need to click sooner after changing heading direction, possibly to update their orientation in relation to bathymetry, it could explain why the more habituated transient animals in Richter et al (2003) study appeared to decrease their time to first click after encountering ww-platforms.

The duration of the first bout of clicks was predicted by the model to increase after encountering whale-watching boats, along with depth and higher blow rates (Models 5 and 6 in Table 2.8 b). However, the duration of first bout was highly variable across individuals and individual differences were predicted to be more than an order of magnitude greater than the average change after encountering whale-watching boats. The ww-vessel covariate remained important when accounting for blow rate, suggesting that whale-watching influenced the duration of the first bout directly, rather than as a consequence of any impact on blow rate. Previous studies have documented small decreases in mean blow interval with whale-watching vessel presence (Gordon et al 1992, Richter et al 2003, 2006). The authors have suggested a stress response. It is possible that some whales attempted to descend further before stopping for their first pause, perhaps to be further away from boat noise or perceived risk.

Note that higher blow rates increased the duration of the first bout of regular clicks, regardless of whale-watching. Higher blow rates and longer surface duration (but not ww-vessel presence) also predicted more regular clicking throughout the dive, which could

mean lower prey encounter rate. These results could either indicate an effect of the research vessel (although see discussion above), or relate to a foraging tactic such as navigating across to a different foraging area while searching for prey opportunistically.

Time to first creak appeared to decrease with number of boats viewing the whale at surface. Along with aspect, however, the model explained very small amount of the small sample (14% in the data of 25 surfacings), and confidence intervals were too large for prediction. Whale-watching vessel presence was not an important predictor of any other creak related variable.

2.5 Conclusions

There was little evidence for changes in parameters related to prey encounter rates that could be conclusively attributed to whale-watching; however, the analysis was constrained by relatively small sample size. The impact of whale-watching flights could not be assessed reliably on any of the acoustic parameters due to small sample size.

There were changes to the initial echolocation behaviour after encountering whale-watching boats: whales delayed their first click and appeared to descend for longer before stopping for their first silence, which was also extended. However, the effects appeared small relative to high individual variability and were within the range of behaviours predicted by their spatio-temporal environment.

2. References

Au W (1993) The sonar of dolphins. Springer-Verlag, New York, US

Cisneros-Montemayor AM, Sumaila UR, Kaschner K, Pauly D (2010) The global potential for whale watching. Mar Policy 34:1273–1278

Clarke M (1996) Cephalopods as prey. III. Cetaceans. Philos Trans R Soc Lond B Biol Sci 351:1053–1065

Corkeron P (2004) Whale watching, iconography, and marine conservation. Conserv Biol 18:847–849

Cranford T (1999) The sperm whale's nose: sexual selection on a grand scale? Mar mammal Sci 15:1133–1157

Davis R, Jaquet N, Gendron D, Markaida U, Bazzino G, Gilly W (2007) Diving behavior of sperm whales in relation to behavior of a major prey species, the jumbo squid, in the Gulf of California, Mexico. Mar Ecol Prog Ser 333:291–302

Drouot V, Gannier A, Goold J (2004) Diving and feeding behaviour of sperm whales (Physeter macrocephalus) in the northwestern Mediterranean Sea. Aquat Mamm 30:419–426

Fristrup K, Harbison G (2002) How do sperm whales catch squids? Mar Mammal Sci 18:42–54

Garner DM (1961) Hydrology of New Zealand Coastal Waters. New Zeal Oceanogr Inst Mem 8:9–65

Gordon J (1987) The behaviour and ecology of sperm whales off Sri Lanka. PhD diss., University of Cambridge

Gordon JCD, Leaper R, Hartley FG, Chappell O (1992) Effects of whale-watching vessels on the surface and underwater acoustic behaviour of sperm whales off Kaikoura, New Zealand. Dep Conserv Sci Res Ser 52

Griffin R (1999) Sperm whale distributions and community ecology associated with a warm-core ring off Georges Bank. Mar Mammal Sci 15:33–51

Heath R (1972) The Southland Current. New Zeal J Mar Freshw Res 6

Higham J, Bejder L, Williams R (2014) Whale-watching: Sustainable Tourism and Ecological Management. Cambridge University Press, Cambridge, UK

Hooker S, Whitehead H, Gowans S (1999) Marine protected area design and the spatial and temporal distribution of cetaceans in a submarine canyon. Conserv Biol 13:592-602

Houtman J (1965) Winter Hydrological Conditions Of Coastal Waters South Of Kaikoura Peninsula. New Zeal J Geol Geophys 8:807–819

Hoyt E (2000) Whale watching 2001: worldwide tourism numbers, expenditures, and expanding socioeconomic benefits. IFAW, Yarmouth Port, US

Jaquet N, Dawson S, Douglas L (2001) Vocal behavior of male sperm whales: Why do they click? J Acoust Soc Am 109:2254

Jaquet N, Dawson S, Slooten E (2000) Seasonal distribution and diving behaviour of male sperm whales off Kaikoura: foraging implications. Can J Zool 1990:407–419

Jaquet N, Whitehead H (1996) Scale-dependent correlation of sperm whale distribution with environmental features and productivity in the South Pacific. Mar Ecol Prog Ser 135:1–9

Johnson M (2008) Echolocation behaviour adapted to prey in foraging Blainville's beaked whale (Mesoplodon densirostris). Proc R Soc B-Biological Sci 275:133–139

Johnson M, Madsen PT, Zimmer WMX, Soto NA de, Tyack PL (2004) Beaked whales echolocate on prey. Proc R Soc B-Biological Sci 271 Suppl:S383–6

Leo FC De, Smith CR, Rowden A a, Bowden D a, Clark MR (2010) Submarine canyons: hotspots of benthic biomass and productivity in the deep sea. Proc Biol Sci 277:2783–92

Lewis KB, Barnes PM (1999) Kaikoura Canyon, New Zealand: active conduit from near-shore sediment zones to trench-axis channel. Mar Geol 162:39–69

Lusseau D (2003) Effects of Tour Boats on the Behavior of Bottlenose Dolphins: Using Markov Chains to Model Anthropogenic Impacts. Conserv Biol 17:1785–1793

Lusseau D, Bejder L (2007) The long-term consequences of short-term responses to disturbance experiences from whalewatching impact assessment. Int J Comp Psychol 20:228–236

Madsen PT, Wahlberg M, Møhl B (2002a) Male sperm whale (Physeter macrocephalus) acoustics in a high-latitude habitat: implications for echolocation and communication. Behav Ecol Sociobiol 53:31–41

Madsen PT, Payne R, Kristiansen NU, Wahlberg M, Kerr I, Møhl B (2002b) Sperm whale sound production studied with ultrasound time/depth-recording tags. J Exp Biol 205:1899–906

Madsen PT, Surlykke A (2013) Functional convergence in bat and toothed whale biosonars. Physiology 28:276–83

Marra G, Wood SN (2011) Practical variable selection for generalized additive models. Comput Stat Data Anal 55:2372–2387

Miller PJO, Johnson MP, Madsen PT, Biassoni N, Quero M, Tyack PL (2009) Using at-sea experiments to study the effects of airguns on the foraging behavior of sperm whales in the Gulf of Mexico. Deep Sea Res Part I Oceanogr Res Pap 56:1168–1181

Miller PJO, Johnson MP, Tyack PL (2004) Sperm whale behaviour indicates the use of echolocation click buzzes "creaks" in prey capture. Proc R Soc London Ser B Biol Sci 271:2239–47

Møhl B, Wahlberg M, Madsen PT, Heerfordt A, Lund A (2003) The monopulsed nature of sperm whale clicks. J Acoust Soc Am 114:1143

Pirotta E, Matthiopoulos J, MacKenzie M, Scott-Hayward L, Rendell L (2011) Modelling sperm whale habitat preference: a novel approach combining transect and follow data. Mar Ecol Prog Ser 436:257–272

Praca E, Gannier A, Das K, Laran S (2009) Modelling the habitat suitability of cetaceans: Example of the sperm whale in the northwestern Mediterranean Sea. Deep Sea Res Part I Oceanogr Res Pap 56:648–657

Rendell L, Whitehead H, Escribano R (2004) Sperm whale habitat use and foraging success off northern Chile: Evidence of ecological links between coastal and pelagic systems. Mar Ecol Prog Ser 275:289–295

Richter CF, Dawson SM, Slooten E (2003) Sperm whale watching off Kaikoura, New Zealand : effects of current activities on surfacing and vocalisation patterns. Sci Conserv 219

Richter C, Dawson S, Slooten E (2006) Impacts of commercial whale watching on male sperm whales at Kaikoura, New Zealand. Mar Mammal Sci 22:46–63

Richter C, Fernandes M, Isojunno S (2011) Effects of vessel traffic on the surface behaviour of individually identified sperm whales foraging in the Kaikoura Canyon. In: Effects of Tourism on the Behaviour of Sperm whales Inhabiting the Kaikoura Canyon.

Thode A, Mellinger DK, Stienessen S, Martinez A, Mullin K (2002) Depth-dependent acoustic features of diving sperm whales (Physeter macrocephalus) in the Gulf of Mexico. J Acoust Soc Am 112:308

Watkins WA (1980) Acoustics and the behavior of sperm whales. In: Animal Sonar Systems (ed. Busnel R). Plenum press, New York, US

Wood SN (2004) Stable and Efficient Multiple Smoothing Parameter Estimation for Generalized Additive Models. J Am Stat Assoc 99:673–686

2. Tables & Figures

Table 2.1 Parameters summarising acoustic behaviour underwater

Short name	Explanation	Min dur
firstRCTime	Time (s) from fluke to first clicks	0
firstRCTrainDur	Duration (s) of first bout of clicks. The bout was considered broken by a silence when the ICI was greater than five times the previous ICI.	0
firstPauseDur	Duration (s) of the first silence, after the first bout of clicks.	0
first5ICI	Initial mean click interval (s). The mean ICI between the first six clicks.	0
last5ICI	Mean ICI (s) at the end of first bout. The mean ICI between the five clicks immediately prior to first silence.	0
ICIChange	Change in ICI (s) during first bout. The difference between initial mean ICI and mean ICI at the end of first bout.	0
firstCTime	Time (s) from fluke to first creak	0
CIFAct	Creak activity – defined as the number of creaks heard divided by the time for which data were analysed after fluke.	30
CIFRate	Creak rate from first creak. The number of creak heard divided by the time for which data were analysed following the first creak.	30
durCI	Proportion of time creaking. Total duration of creak intervals divided by analysis time.	30
medICIRC	Median ICI (s)	20
propMedICIRC	Proportion of clicks within 0.1 seconds of the median ICI	20
durPauseRC	Proportion of time spent silent. Total duration of gaps in regular clicking where no creaks were detected, divided by analysis time.	20

The first column lists the names used as response variables in the model. “Min dur” is the minimum analysis duration since fluke-up and “ICI” the inter-click-interval.

Table 2.2 Covariate data

	Variable	Explanation
Spatial	*x, y*	The location of the whale's fluke-up was calculated using the research vessel's GPS position and the range and bearing estimates to the whale at fluke-up, corrected with ~22 degrees of positive magnetic variation.
	depth	Sea bottom depth at fluke-up (m). Depth values were extracted by overlaying the fluke-up coordinates on a 250 m resolution depth made available by NIWA online at http://www.niwa.co.nz/our-science/oceans/bathymetry/download?sid=415
	slope, aspect	Slope (0-90°) and aspect (180-180°) of the sea bottom at fluke-up. Slope and aspect were computed from the depth surface using a 3-by-3 window of grid cells in Manifold GIS software (www.manifold.net).
Temporal	*Year*	Year 2009-2011
	month	Month 1-12
Surface behaviour	*surfDur*	Surface Duration (min). Only for surfacings where the first or early blows and fluke-up were observed
	blowRate	Number of blows per minute.
	surfSpeed	Speed over ground during surfacing (m/s). Measured between the first and last observation of the whale at surface.
	HChange	Change in heading during surfacing (°). The difference in estimated heading between the beginning and end of surfacing. When the heading was estimated at ranges > 600 metres, the travelling heading of the preceding/following dive was computed between the fluke-up and surface location, where available.
Whale-watching	*boats*	Number of ww-boats
	copters	Presence of ww-helicopters
	planes	Presence of ww-aeroplanes.
	vessels	Types of ww-vessels present. 0 – no ww-vessels, 1 – only ww-boats present, 2 – only ww-helicopters or aeroplanes present 3 – both ww-boats and ww-flights present

Table 2.3 Sample size

		All analysed data	Complete dive cycles
No ww-vessels	N	40	23
	Follows / identified	16 / 9	14/9
Only ww-boats present	N	21	13
	Follows / identified	16 / 9	10/5
Only ww-flights present	N	4	2
	Follows / identified	4 / 3	2/2
Both ww-boats and ww-flights present	N	11	8
	Follows / identified	6 / 3	5/3
Total	N	76	46
	Follows / identified	22 / 11	18/10

Number of surfacings (N), number of follow occasions (Follows) and reliably identified whales broken down by presence of whale-watching vessels. Left column shows all data, and complete dive cycles on the right.

Table 2.4 Descriptive mean and range statistics for each acoustic parameter

	First regular click train					Creaking						
	first RCTime (s)	*first RCTrain Dur* (s)	*first Pause Dur* (s)	*first 5ICI* (s)	*ICI Change (s)*	*first CTime (min)*	*CIF Act* (-h)	*CIF Rate* (-h)	*dur CI* %	*ICI Med* (s)	*propMed ICIRC* %	*dur PauseRC* %
No ww-vessels												
N	40/16	40/16	40/16	40/16	40/16	32/15	23/14	23/14	23/14	30/14	30/14	30/14
Individual ids	9	9	9	9	9	9	9	9	9	9	9	9
Mean	8.71	102.42	8.87	1.13	0.00	13.17	8.02	11.70	9.6	1.05	21.0	14.1
Sd	7.75	82.11	4.65	0.26	0.32	6.46	6.65	8.35	9.3	0.17	15.3	7.2
Min	0.13	6.67	1.13	0.72	-0.80	2.53	0.00	0.00	0.0	0.72	2.7	4.4
Max	30.79	389.82	27.77	1.67	0.71	29.29	21.25	27.69	38.7	1.37	69.0	34.9
Only ww-boats												
N	21/16	21/16	21/16	21/16	21/16	11/9	14/11	14/11	14/11	13/11	13/11	14/11
Individual ids	9	9	9	9	9	7	6	6	6	7	7	7
Mean	8.19	125.02	11.20	1.10	0.13	10.48	6.18	8.51	9.0	1.07	25.6	12.8
Sd	7.45	98.09	5.40	0.31	0.53	5.40	6.26	7.76	14.4	0.24	14.7	6.0
Min	0.07	18.13	6.86	0.52	-0.70	4.86	0.00	0.00	0.0	0.84	7.3	4.7
Max	22.23	335.45	29.51	1.54	1.20	19.06	20.09	24.29	50.7	1.53	51.4	27.1
Only ww-flights												
N	4/4	4/4	4/4	4/4	4/4	2/2	2/2	2/2	2/2	3/3	3/3	3/3
Individual ids	3	3	3	3	3	2	2	2	2	3	3	3
Mean	10.25	49.89	5.00	0.93	-0.06	6.22	2.19	2.52	5.4	0.94	33.2	7.9
Sd	7.01	39.12	3.39	0.67	0.43	-	-	-	-	-	-	-
Min	0.00	0.00	0.00	0.00	-0.68	5.35	0.00	0.00	0.0	0.79	26.4	0.0
Max	15.24	95.10	7.45	1.54	0.30	7.09	4.39	5.04	10.8	1.16	40.8	16.4

Parameters measured during diving (all analysed data included) are broken down by the presence of whale-watching vessels in the preceding surfacing. Both number of surfacings and number of follows are given as sample size (N).

Table 2.5 Descriptive mean and range statistics for parameters measured for complete dive cycles

	Dive duration (min)	Speed over ground (kt)	Dir change (deg)	Creak activity *CIFAct* (-h)	Creak rate *CIFRate (-h)*	Time spent creaking *durCI (%)*	Median regular ICI *ICIMed (s)*	Clicks at median ICI *propMed ICIRC (%)*	Time spent Silent *durPauseRC (%)*
No ww-vessels									
N	23/14	23/14	23/14	19/14	19/14	19/14	21/14	21 /14	21/14
Individual ids	9	9	9	9	9	9	9	9	9
Mean	35.30	0.66	55.19	6.79	10.30	9.0	1.06	20.4	14.2
Sd	6.97	0.28	44.18	5.44	6.87	9.6	0.17	16.7	7.9
Min	19.77	0.10	1.79	1.35	1.97	0.6	0.72	2.7	4.4
Max	49.62	1.20	154.95	18.90	23.92	38.7	1.37	69.0	34.9
Only ww-boats									
N	13/10	13/10	11/9	10/9	10/9	10/9	11/9	11/9	12/9
Individual ids	5	5	5	5	5	5	5	5	5
Mean	34.01	0.80	54.66	6.58	8.32	11.2	1.09	25.1	11.7
Sd	10.23	0.44	52.28	7.18	8.57	16.7	0.26	13.7	4.7
Min	11.82	0.30	4.82	0.00	0.00	0.0	0.84	7.3	4.7
Max	44.77	2.01	159.98	20.09	24.29	50.7	1.53	51.4	19.3
Only ww-flights									
N	2/2	1/1	2/2	1/1	1/1	1/1	1/1	1/1	1/1
Individual ids	2	1	2	1	1	1	1	1	1
Mean	27.58	0.27	30.51	4.39	5.04	10.8	1.16	26.4	16.4
Min	0.48		27.88						
Max	54.68		33.14						

Parameters broken down by the presence of whale-watching vessels in the preceding surfacing. Both number of surfacings and number of follows are given as sample size (N), e.g. 23/14: 23 surfacings, 14 follow occasions.

Table 2.6 Wilcoxon matched-pairs signed-rank tests

	Type 1: No-boat encounter vs. subsequent boat encounter					**Type 2:** Boat encounter vs. subsequent no-boat encounter					**Type 3:** Previous encounter vs. subsequent encounter				
	N	V	p	sign	median diff	N	V	p	sign	median diff	N	V	p	sign	median diff
firstRCTime	8	14	0.641	+	3.53	11	37	0.765	-	-0.01	3	0	0.250	+	2.79
firstRCTrainDur	8	10	0.313	+	10.75	11	47	0.240	-	-39.50	3	0	0.250	+	35.76
firstPauseDur	8	0	0.008	+	2.53	11	55	0.054	-	-3.89	3	2	0.750	+	0.08
first5ICI	8	11	0.383	+	0.19	11	30	0.831	+	0.14	3	2	0.750	+	0.53
last5ICI	8	12	0.461	+	0.23	11	42	0.465	-	-0.02	3	2	0.750	+	0.18
ICIChange	8	15	0.742	+	0.07	11	41	0.520	-	-0.05	3	3	1.000	+	0.10
firstCTime	4	4	0.875	+	2.43	3	0	0.250	+	3.32	1				
CIFAct	3	1	0.500	+	4.74	4	3	0.625	+	1.65	0				
CIFRate	3	1	0.500	+	6.25	4	1	0.250	+	4.58	0				
durCI	3	1	0.500	+	4.73	4	2	0.375	+	3.76	0				
medICIRC	6	15	0.438	-	-0.05	4	8	0.375	-	-0.14	1				
propMedICIRC	6	16	0.313	-	-8.49	4	6	0.875	-	-2.91	1				
durPauseRC	6	10	1.000	+	2.34	4	3	0.625	+	3.46	1				

(Table 2.6 continued on next page)

	Type 1: No-flight encounter vs. subsequent flight encounter					**Type 2:** Flight encounter vs. subsequent no-flight encounter				
	N	V	p	sign	median diff	N	V	p	sign	median diff
firstRCTime	5	12	0.313	-	-18.58	3	0	0.250	+	3.24
firstRCTrainDur	5	10	0.625	-	-10.42	3	0	0.250	+	22.10
firstPauseDur	5	10	0.625	-	-0.32	3	3	1.000	+	-0.75
first5ICI	5	9	0.813	-	0.05	3	0	0.250	+	0.31
last5ICI	5	6	0.813	+	0.16	3	0	0.250	+	0.19
ICIChange	5	5	0.625	+	0.08	3	4	0.750	-	-0.03
firstCTime	2	3	0.500	-	-8.07	0				
CIFAct	3	1	1.000	+	0.00	0				
CIFRate	3	1	1.000	+	0.00	0				
durCI	3	1	1.000	+	0.00	0				
medICIRC	3	5	0.500	-	-0.15	0				
propMedICIRC	3	2	0.750	+	0.01	0				
durPauseRC	4	0	0.125	+	0.10	0				

Number of pairs tested, test statistic (sum of negative signed ranks) and p values are given in columns N, V and p, respectively. 'Sign' is the sign of the maximum total rank, and can be interpreted as increase or decrease in the values. Also the median pair-wise difference is given for each acoustic parameter.

Table 2.7 Distribution, link function, and fitted values of all models

Model no	Response variable	Model type	Distribution	Link	Var	N	Rsq %	Autocorr (Phi)	Min Resid	Max Resid	Min Fitted	Max Fitted	Min Observed	Max Observed
1	firstRCTime	null	Tweedie(1.1)	mu^0.1		33	26.98	0.67	-4.01	3.51	1.50	25.29	0.13	30.79
2		ww	Tweedie(1.1)	mu^0.1		59	37.12	0.49	-3.97	3.54	0.43	26.21	0.13	30.79
3		all	Tweedie(1.1)	mu^0.1		56	43.30	0.48	-3.88	3.20	0.68	25.90	0.13	30.79
4	firstRCTrainDur	null	quasi	log	mu	33	56.44	-0.53	-12.90	9.90	32.03	384.35	12.22	389.82
5		ww	quasi	log	mu	59	51.87	-0.29	-8.65	11.99	23.33	273.19	12.22	389.82
6		all	quasi	log	mu	56	51.52	-0.75	-11.53	12.36	32.04	345.62	12.22	389.82
7	firstPauseDur	null	quasi	inverse	mu	33	0.00	-0.11	-3.20	5.20	8.54	8.54	1.13	27.77
8		ww	quasi	inverse	mu	59	7.40	-0.19	-3.26	5.54	4.91	22.02	1.13	29.51
9		all	quasi	inverse	mu	56	8.28	-0.20	-3.27	5.54	4.90	22.28	1.13	29.51
10	first5ICI	null	gaussian	identity		33	60.58	0.00	-0.38	0.29	0.80	1.39	0.73	1.65
11		ww	gaussian	identity		59	75.81	-0.14	-0.25	0.27	0.72	1.55	0.52	1.65
12		all	gaussian	identity		56	75.21	-0.13	-0.25	0.27	0.73	1.55	0.52	1.65
13	last5ICI	null	Tweedie(1.1)	mu^0.1		33	16.91	-0.17	-0.46	0.41	0.94	1.32	0.66	1.64
14		ww	Tweedie(1.1)	mu^0.1		59	18.15	-0.29	-0.39	0.36	0.87	1.55	0.66	1.66
15		all	Tweedie(1.1)	mu^0.1		56	57.32	-0.08	-0.26	0.30	0.72	1.69	0.66	1.66
16	ICIChange	null	gaussian	identity		33	25.71	-0.33	-0.52	0.58	-0.28	0.30	-0.80	0.71
17		ww	gaussian	identity		59	40.01	0.06	-0.63	0.54	-0.49	0.80	-0.80	0.92
18	firstCTime		Gamma	log		25	13.68	0.78	-0.88	0.92	7.19	16.81	4.93	29.29
19	CIFAct		quasi	inverse	mu	30	18.46	0.59	-4.61	2.52	5.53	11.51	0.00	20.09
20	CIFRate		quasipoisson	log		30	0.00	0.50	-4.59	3.62	10.52	10.52	0.00	24.29
21	durCI		quasi	log	mu	30	45.04	0.58	-0.50	0.48	0.02	0.60	0.00	0.62
22	medICIRC		gaussian	log		38	85.57	-0.44	-0.13	0.13	0.68	1.52	0.61	1.53
23	propMedICIRC		quasi	identity		41	45.15	-0.80	-0.20	0.21	-0.01	0.57	0.03	0.58
24	durPauseRC		quasi	log	mu	42	34.69	0.39	-0.23	0.26	0.09	0.23	0.04	0.26

Table 2.8 a All models with the retained covariates

	response	model	depth			slope			aspect			year			month		
			EDF	F	p	EDF	F	p	EDF	F	p	EDF	F	p	EDF	F	p
1	firstRCTime	null	NA	NA	NA	NA	NA	NA	0.861	5.280	0.034	1.199	6.530	0.012	0.906	8.499	0.008
2		ww	NA	NA	NA	NA	NA	NA	0.938	7.898	0.008	NA	NA	NA	NA	NA	NA
3		all	NA	NA	NA	0.517	0.865	0.286	0.946	8.673	0.006	NA	NA	NA	0.683	2.345	0.138
4	firstRCTrainDur	null	NA	NA	NA	NA	NA	NA	NA	NA	NA	1.710	22.365	0.000	NA	NA	NA
5		boat	2.440	5.767	0.004	NA	NA	NA	0.830	3.807	0.065	NA	NA	NA	NA	NA	NA
6		all	1.069	22.664	0.000	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
7	firstPauseDur	null	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
8		ww	0.852	3.554	0.072	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
9		all	0.849	3.623	0.070	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
10	first5ICI	null	1.046	18.123	0.000	NA	NA	NA	1.009	14.563	0.001	0.749	2.470	0.131	NA	NA	NA
11		ww	1.564	75.353	0.000	0.920	5.950	0.021	NA	NA	NA	NA	NA	NA	0.964	22.633	0.000
12		all	1.507	69.427	0.000	0.905	5.448	0.028	NA	NA	NA	NA	NA	NA	0.962	21.772	0.000
13	last5ICI	null	NA	NA	NA	NA	NA	NA	0.939	7.642	0.011	NA	NA	NA	NA	NA	NA
14		ww	1.930	1.840	0.173	NA	NA	NA	0.760	2.537	0.124	NA	NA	NA	NA	NA	NA
15		all	3.777	7.298	0.000	NA	NA	NA	0.996	13.153	0.001	NA	NA	NA	NA	NA	NA
16	ICIChange	null	0.995	13.655	0.001	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
17		ww	1.110	15.972	0.000	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
18	firstCTime								0.632	1.560	0.208						
19	CIFAct								0.859	4.788	0.042				NA	NA	NA
20	CIFRate								NA	NA	NA						
21	durCI		0.941	4.781	0.043	NA	NA	NA									
22	medICIRC		NA	NA	NA										0.884	6.171	0.024
23	propMedICIRC		1.034	15.280	0.001	NA	NA	NA	1.041	22.878	0.000				NA	NA	NA
24	durPauseRC		NA	NA	NA	1.091	29.952	0.000	NA	NA	NA						

NA:s show covariates that were fitted in each full model, but were then excluded by shrinkage model selection. Columns EDF/df, F, and p values show (estimated) degrees of freedom, Wald test statistic and p-value, respectively.

Table 2.8 b All models with the retained covariates

	surfDur			blowRate			surfSpeed			HChange			Boats			as.factor(vessel)			as.factor(individual)			
	EDF	F	p	EDF	F	p	EDF	F	p	EDF	F	p	EDF	F	p	df	F	p	df	F	p	no sig levels
1	NA	NA	NA	NA	NA	NA	NA	NA	NA	0.749	2.716	0.117										
2													NA	NA	NA	3	2.331	0.088	11	2.779	0.008	5
3	NA	NA	NA	NA	NA	NA	NA	NA	NA	0.889	4.860	0.037	NA	NA	NA	3	3.538	0.023	10	3.586	0.002	8
4	NA	NA	NA	1.167	33.842	0.000	NA	NA	NA	NA	NA	NA										
5													NA	NA	NA	3	2.865	0.048	11	7.816	0.000	4
6	NA	NA	NA	1.077	12.364	0.001	NA	NA	NA	NA	NA	NA	NA	NA	NA	3	3.217	0.033	10	17.143	0.000	5
7	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA										
8													NA	NA	NA	3	1.128	0.348	11	1.327	0.243	1
9	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	3	1.080	0.368	10	1.416	0.208	0
10	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA										
11													NA	NA	NA	3	0.366	0.778	11	5.520	0.000	7
12	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	3	0.346	0.792	10	5.730	0.000	6
13	NA	NA	NA	NA	NA	NA	NA	NA	NA	0.559	1.185	0.244										
14													NA	NA	NA	3	1.783	0.165	11	2.825	0.008	1
15	2.015	5.207	0.011	NA	NA	NA	NA	NA	NA	2.333	4.377	0.016	NA	NA	NA	3	2.973	0.046	10	3.630	0.002	2
16	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA										
17													NA	NA	NA	3	0.279	0.840	11	0.675	0.754	0
18							NA	NA	NA				0.909	5.575	0.030							
19																						
20																						
21																			9	0.831	0.597	2
22	NA	NA	NA				NA	NA	NA	0.816	2.691	0.119				3	2.466	0.089	10	24.453	0.000	5
23	1.062	8.930	0.006	0.934	6.622	0.018													11	15.379	0.000	8
24																						0

NA:s show covariates that were fitted in each full model, but were excluded by shrinkage. Columns EDF/df, F, and p values show (estimated) degrees of freedom, Wald test statistic and p-value, respectively.

Table 2.9 Coefficient estimates for each vessel factor

response	modelNo	Vessel factor	Estimate	Std Error	t-value	p-value
Time from fluke to first clicks *firstRCTime*	2	Intecept	1.115	0.045	24.764	0.000
		ww-boats	0.037	0.020	1.875	0.068
		ww-flights	0.085	0.061	1.393	0.171
		both	-0.001	0.047	-0.015	0.988
	3	Intecept	1.051	0.054	19.309	0.000
		ww-boats	0.048	0.019	2.523	0.016
		ww-flights	0.135	0.060	2.249	0.030
		both	0.051	0.050	1.019	0.314
Duration of first bout of clicks *firstRCTrainDur*	5	Intecept	3.813	0.237	16.121	0.000
		ww-boats	0.520	0.178	2.920	0.006
		ww-flights	0.423	0.465	0.910	0.368
		both	0.497	0.340	1.462	0.151
	6	Intecept	4.163	0.189	21.972	0.000
		ww-boats	0.241	0.137	1.753	0.087
		ww-flights	0.337	0.343	0.985	0.331
		both	0.766	0.256	2.991	0.005
Duration of first silence *firstPauseDur*	8	Intecept	0.118	0.021	5.625	0.000
		ww-boats	-0.026	0.016	-1.688	0.099
		ww-flights	0.000	0.057	0.006	0.996
		both	0.003	0.038	0.086	0.932
	9	Intecept	0.118	0.021	5.533	0.000
		ww-boats	-0.026	0.016	-1.649	0.107
		ww-flights	0.000	0.059	0.008	0.994
	9	both	0.004	0.039	0.094	0.925
Initial mean click interval *first5ICI*	11	Intecept	1.406	0.093	15.085	0.000
		ww-boats	0.001	0.048	0.018	0.985
		ww-flights	0.013	0.114	0.114	0.910
		both	-0.071	0.092	-0.770	0.446
	12	Intecept	1.402	0.096	14.620	0.000
		ww-boats	0.001	0.049	0.026	0.979
		ww-flights	0.014	0.117	0.118	0.907
		both	-0.070	0.094	-0.742	0.463
Mean ICI at the end of first bout *last5ICI*	14	Intecept	1.007	0.009	113.861	0.000
		ww-boats	0.003	0.007	0.477	0.636
		ww-flights	0.019	0.017	1.151	0.256
		both	0.029	0.013	2.264	0.029
	15	Intecept	1.001	0.008	121.586	0.000
		ww-boats	0.005	0.006	0.906	0.372
		ww-flights	0.027	0.013	2.042	0.049
		both	0.028	0.010	2.870	0.007
Change in ICI during first bout *ICIChange*	17	Intecept	-0.026	0.130	-0.203	0.840
		ww-boats	0.028	0.100	0.278	0.782
		ww-flights	0.137	0.238	0.575	0.569
		both	0.159	0.175	0.908	0.369
Median ICI *medICIRC*	22	Intecept	-0.121	0.055	-2.224	0.037
		ww-boats	0.069	0.035	1.988	0.059
		ww-flights	0.140	0.064	2.169	0.041
		both	0.117	0.058	2.013	0.056

Yellow highlights significant effects (at 5%-level)

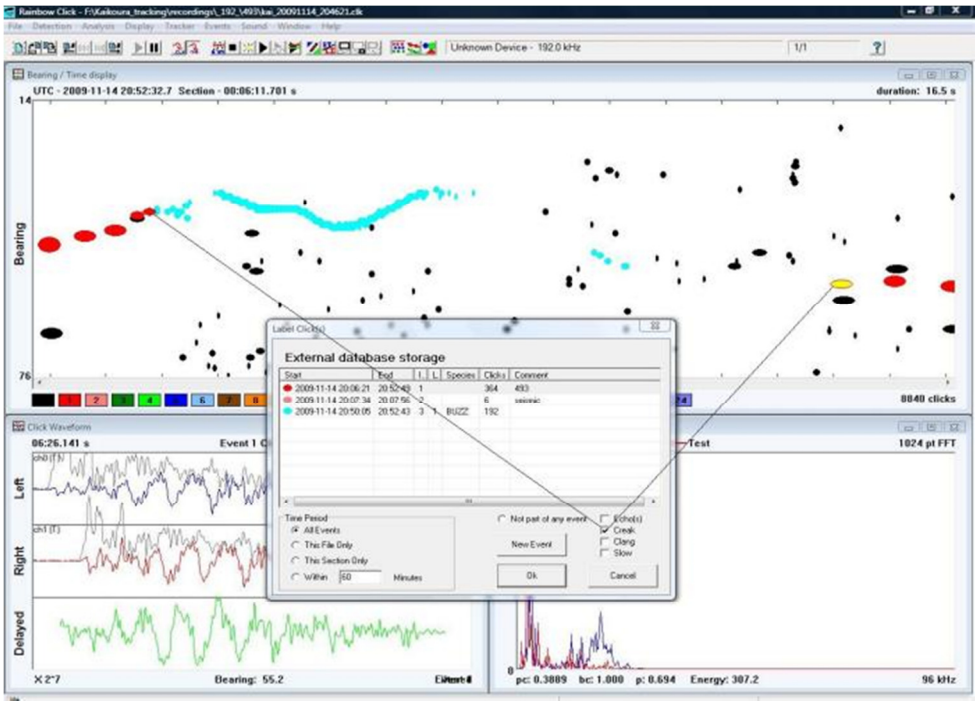


Figure 2.1 Rainbow click bearing display showing marking of a creak produced by the followed whale

X-axis: time, Y-axis: bearing from the boat to the whale, black dots: automated click detections. The regular clicks produced by the followed focal whale have been marked red and creak clicks blue. Dot size is proportional to their received level of each click. The last and first regular click in either side of the creak were marked to show the gap in regular clicking that contained the creak, i.e. 'creak interval'.

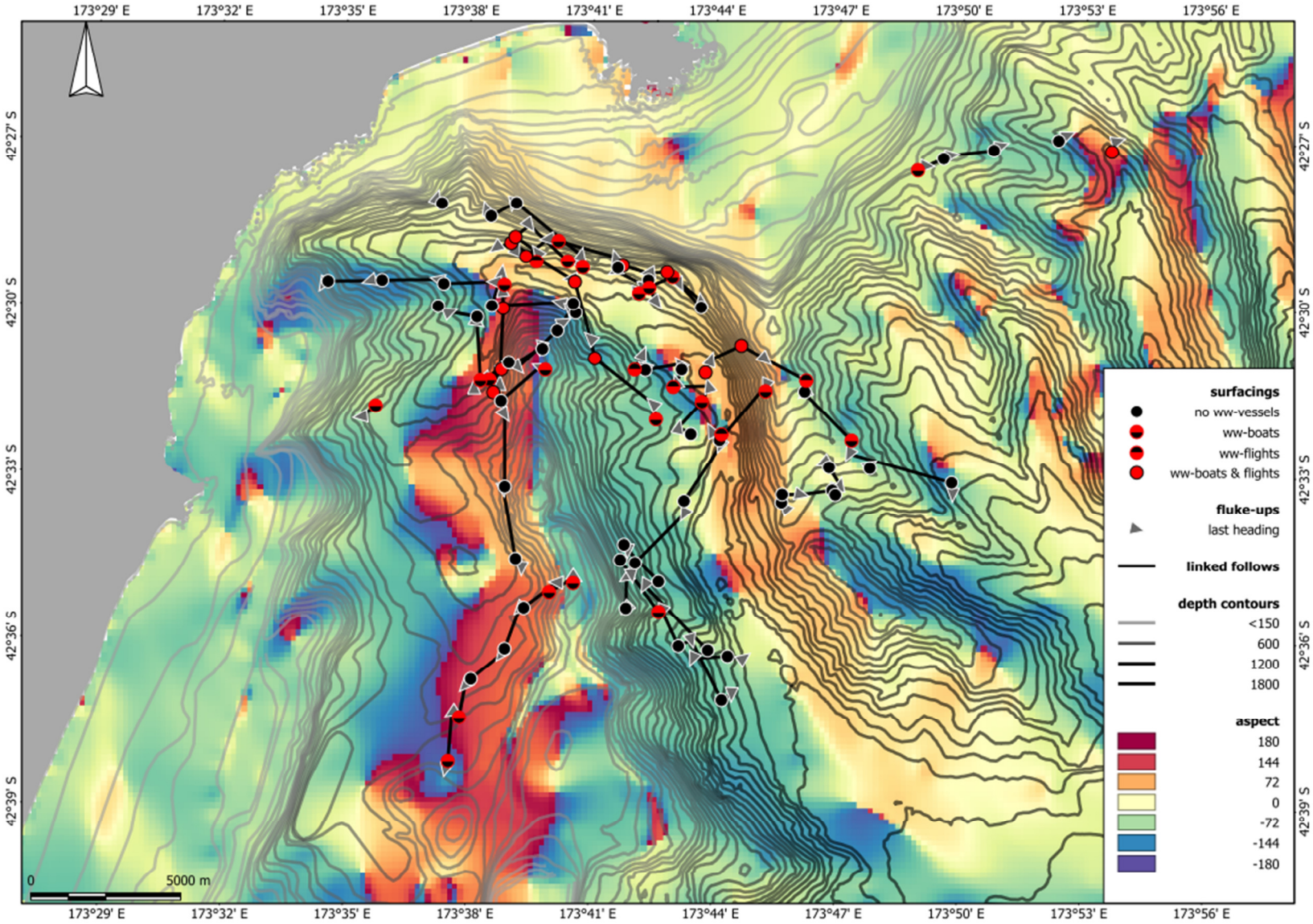


Figure 2.2 Map of surfacings and focal follows for which acoustic data was extracted

Black lines link surfacing locations that were confirmed to be those of the same whale.

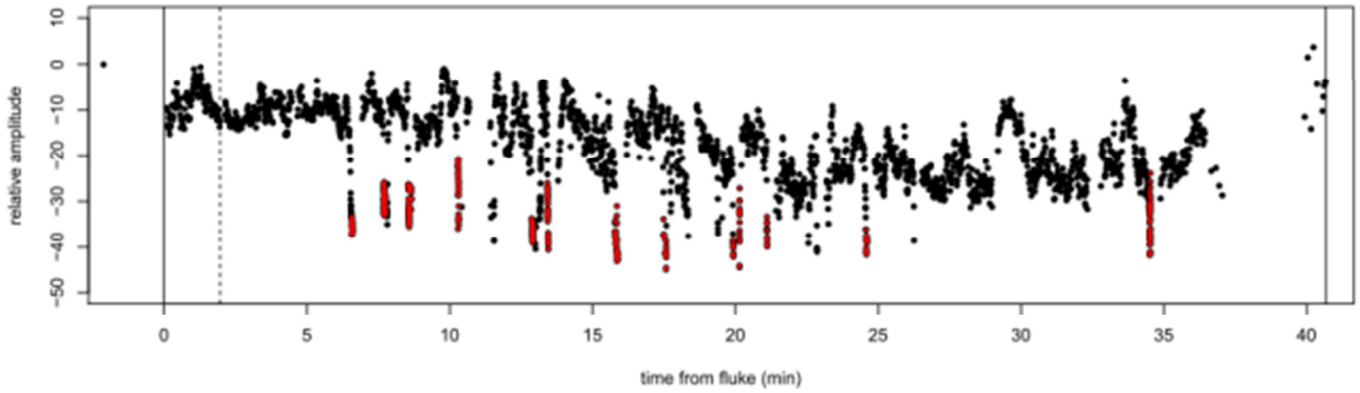


Figure 2.3 Example of an extracted dive cycle

Black dots: regular clicks, red dots: detected creak clicks, solid lines: fluke and surfacing time, dashed line: the start time of the first pause. Amplitude is shown in dB relative to that of the first click of the dive cycle (in this case, at the surface). Note the slow clicks produced at the end of the dive. This whale was viewed by 3 whale-watching boats and one aeroplane.

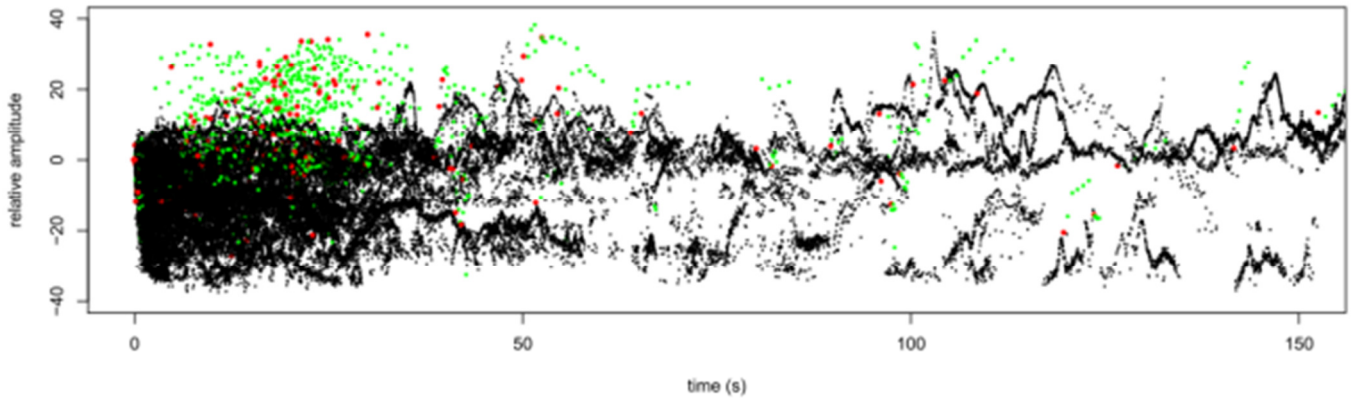


Figure 2.4 Detected creak trains (black) overlaid as a function of time (s) since last regular click

Red dots: the last and first regular clicks (hence marking the start and end of a 'creak interval'), green dots: 5 regular clicks following each creak. Amplitude is given relative to the last regular click before the creak train.

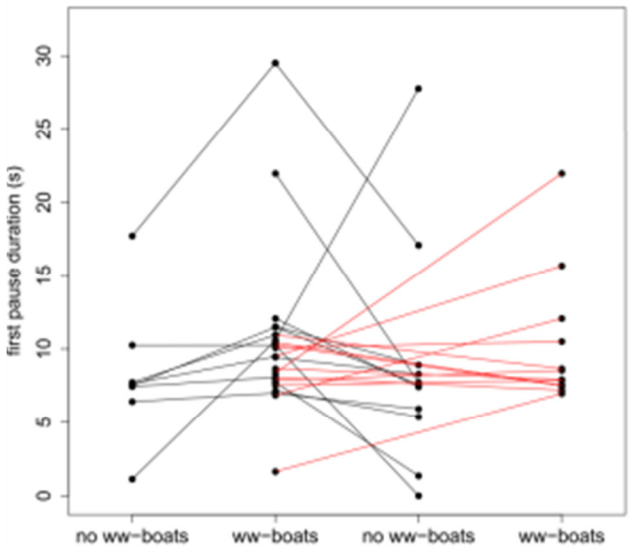


Figure 2.5 First pause duration firstPauseDur (s) *before, during and after exposure to ww-boats*

Red lines: adjacent dives that were both exposed.

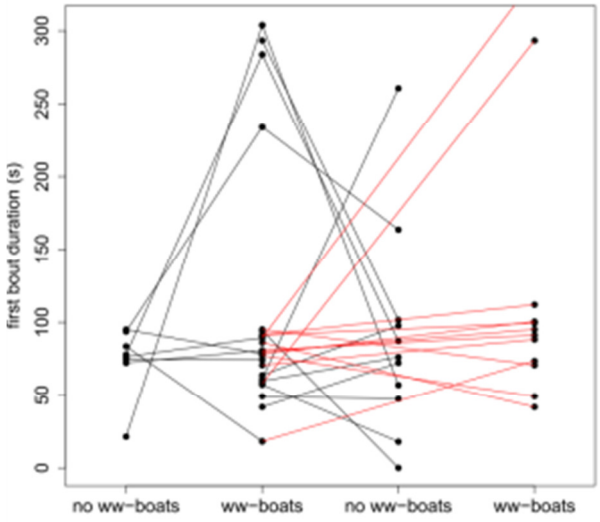


Figure 2.6 First bout duration firstRCTrainDur (s) *before, during and after exposure to ww-boats*

Red lines: adjacent dives that were both exposed.

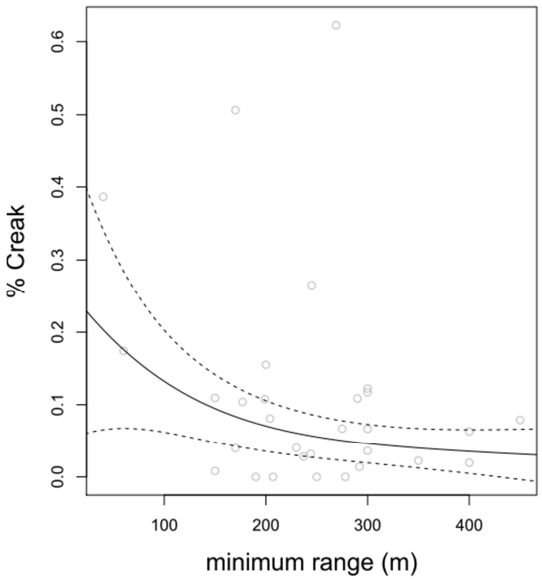


Figure 2.7 Predicted proportion of time in creak intervals durCI (%)

Predictions are shown as a function of minimum range to the research vessel, given observed mean value for depth. Dashed lines: confidence intervals assuming t-distribution, grey circles: data.

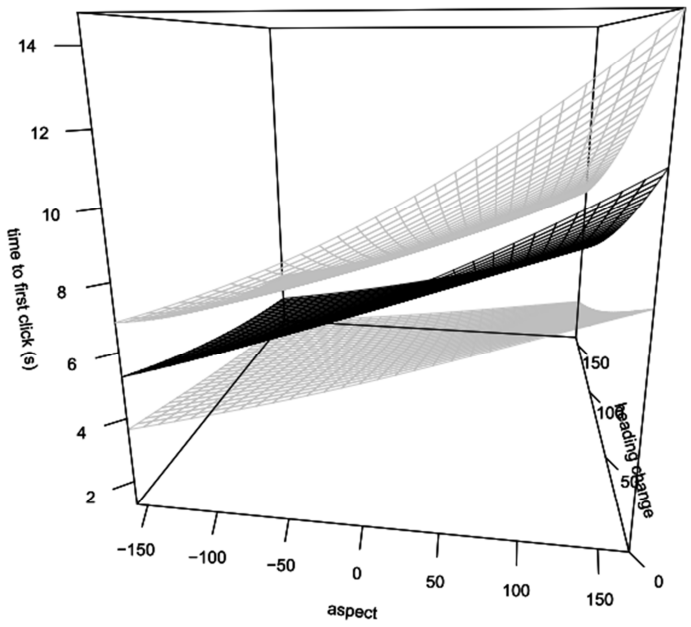


Figure 2.8 Predictions for time to first click firstRCTime (s)

Predictions are given as a function of both aspect of sea bottom and heading change (Model 3). Black: mean, grey: +/- standard errors.

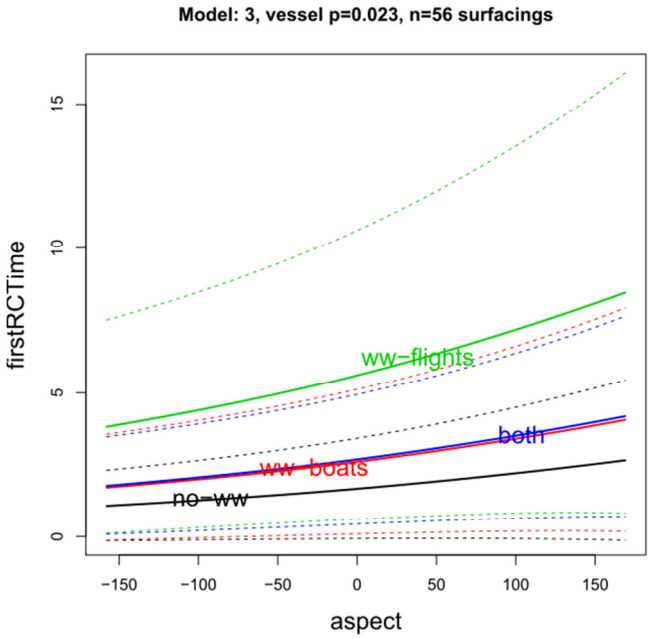


Figure 2.9 Predicted time to first click firstRCTime (s) *after encounters with different ww-vessels*

Predictions are shown as a function of sea bottom aspect at fluking, given mean values for all other covariates and individual factor fixed to its intercept under Model 3 (Table 8 a). Dashed lines: 95% confidence intervals assuming t-distribution.

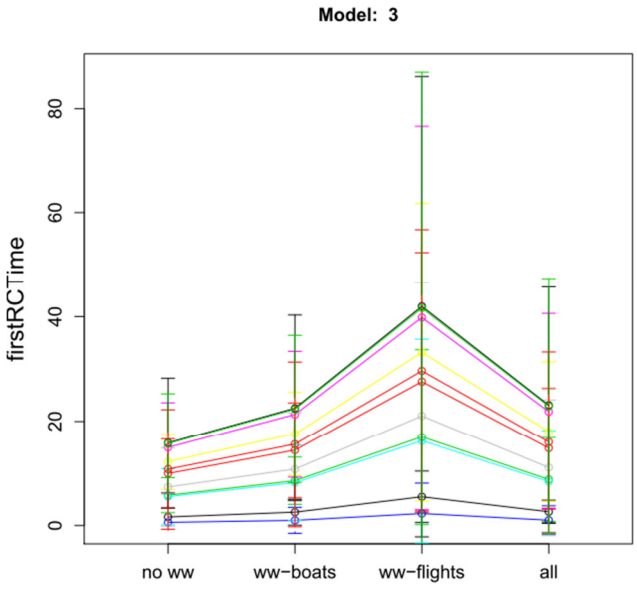


Figure 2.10 Predicted time to first click firstRCTime (s) *by individual after encounters with different ww-vessels*

Predictions are given mean values for all other covariates under Model 3 (Table 8). Individuals are shown in different colours. Crossbars show 95% confidence intervals assuming t-distribution.

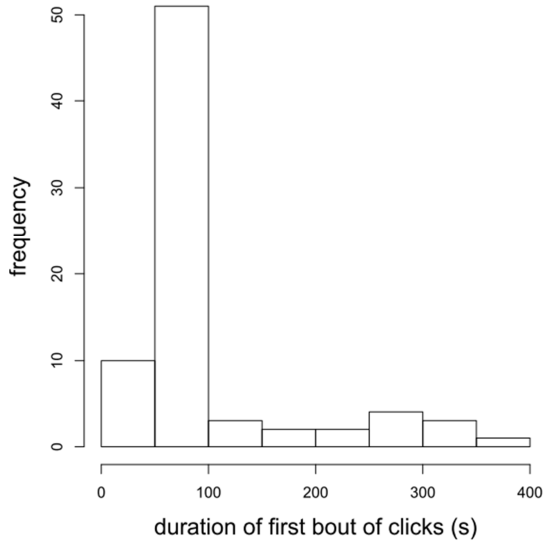


Figure 2.11 Distribution of the duration of first bout of clicks firstRCTrainDur

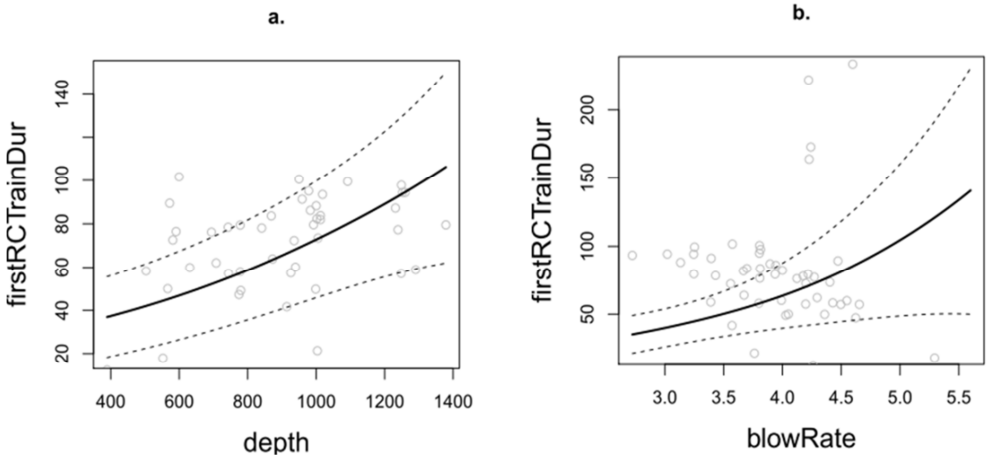


Figure 2.12 Predicted duration (s) of first bout of regular clicks firstRCTrainDur

Predictions are shown as function of fluking depth (a) and blow rate (blows per minute) (b) given no whale-watching vessels, individual fixed to its intercept and all other covariates fixed to their means (mean depth 885 m) under Model 6. Dashed lines: 95% confidence interval assuming t-distribution, grey circles: data.

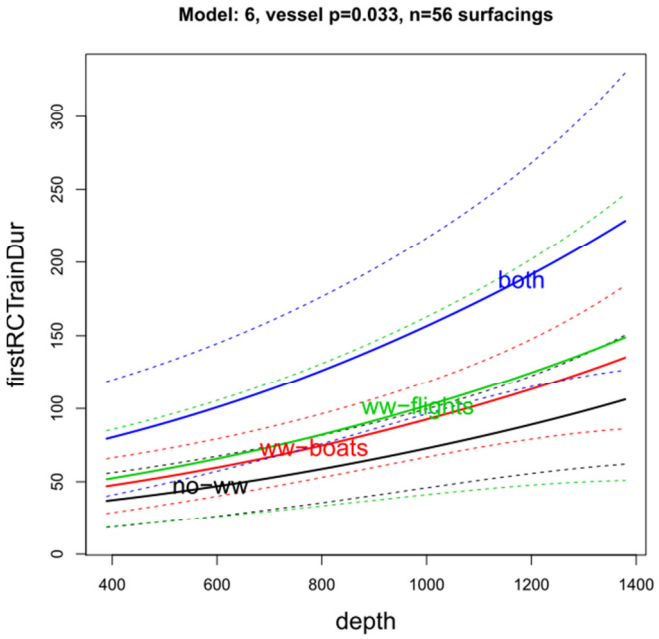


Figure 2.13 Predicted duration of first clicking bout (s) in different exposure conditions

Predictions are given as a function of fluking depth (m), mean values for all other covariates and individual factor fixed to its intercept under Model 6 (Table 8 a). Dashed lines: 95% confidence intervals assuming t-distribution.

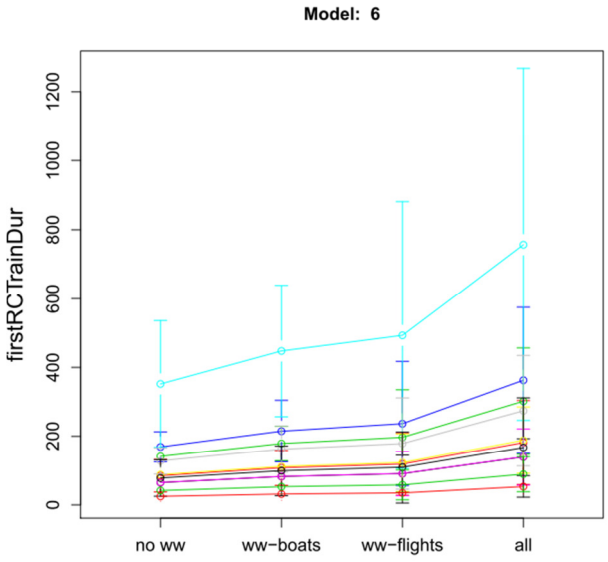


Figure 2.14 Predicted individual duration of first bout of clicks after different whale-watching encounters

Predictions are given mean values for all other covariates under Model 6 (Table 8). Individuals are shown in different colours. Crossbars: 95% confidence intervals assuming t-distribution.

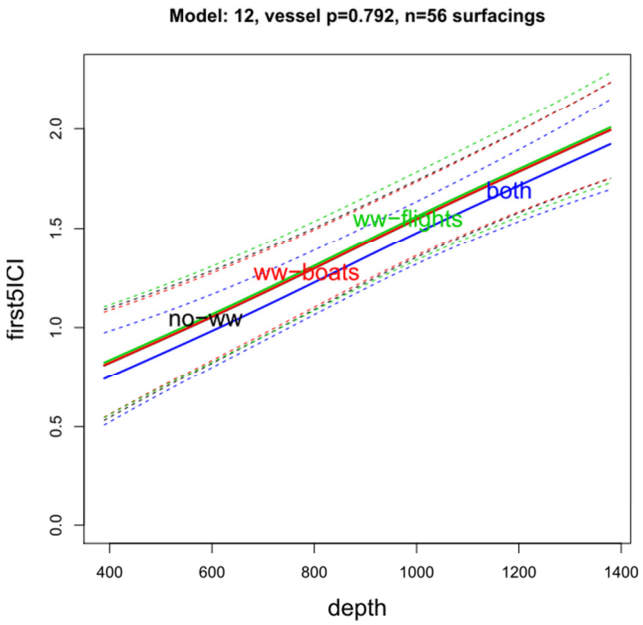


Figure 2.15 Predicted mean ICI (s) after encounters with different ww-vessels

Predictions are given as a function of depth (m), mean values for all other covariates and individual factor fixed to its intercept under Model 12 (Table 8 a, Table 9). Dashed lines: 95% confidence intervals assuming t-distribution.

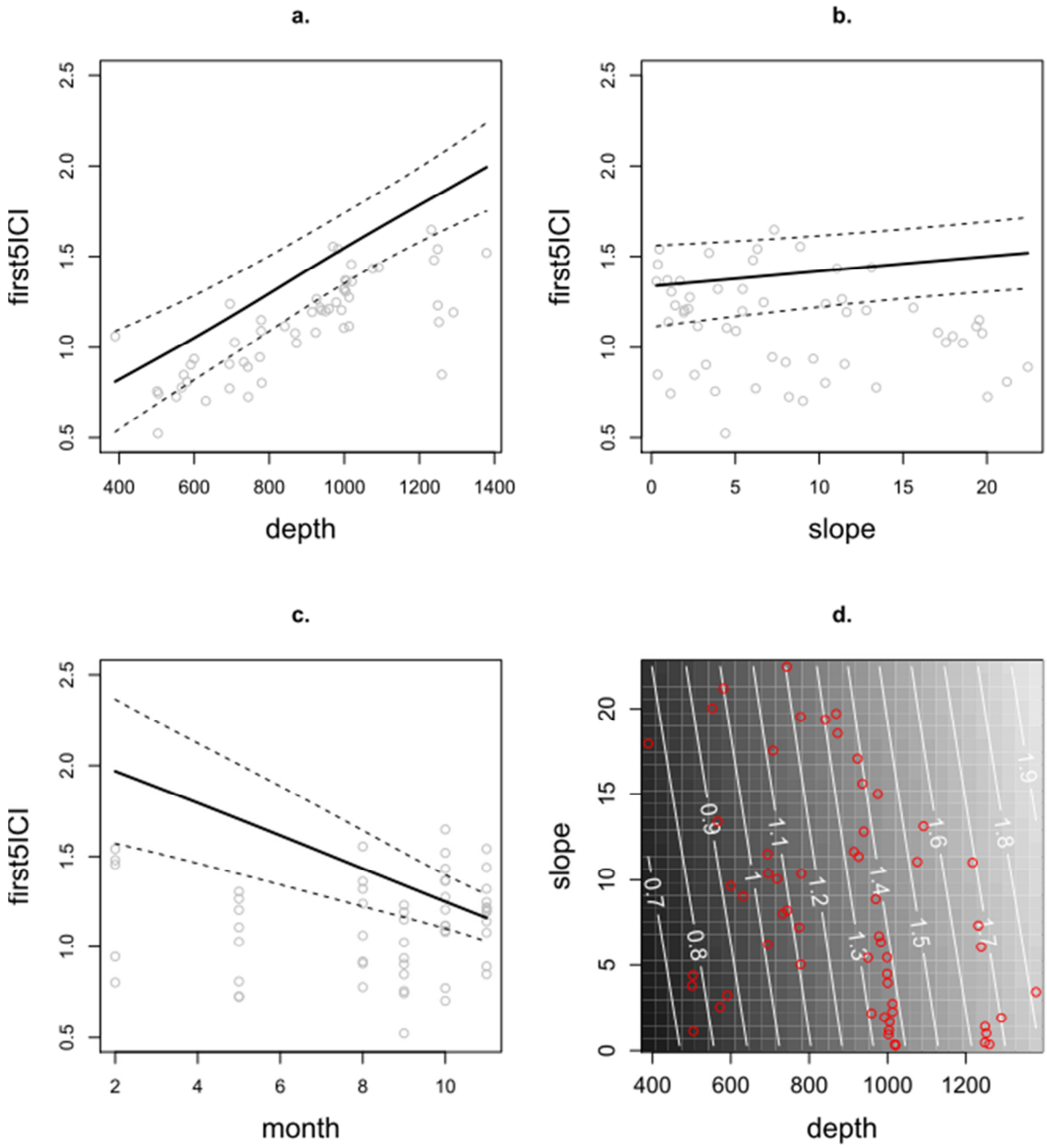


Figure 2.16 Predicted mean initial ICI first5ICI *(s)*

Predictions are given as a function of depth (a), slope (b) and month (c), with no whale-watching vessels in the previous surfacing, individual fixed to its intercept and all other covariates fixed to their means (mean depth 885 m) under Model 15. Dashed lines: 95% confidence interval assuming t-distribution. In plot (d), predicted mean initial ICI is given as a function of both depth and slope with all other covariate values fixed. Grey and red circles: data.

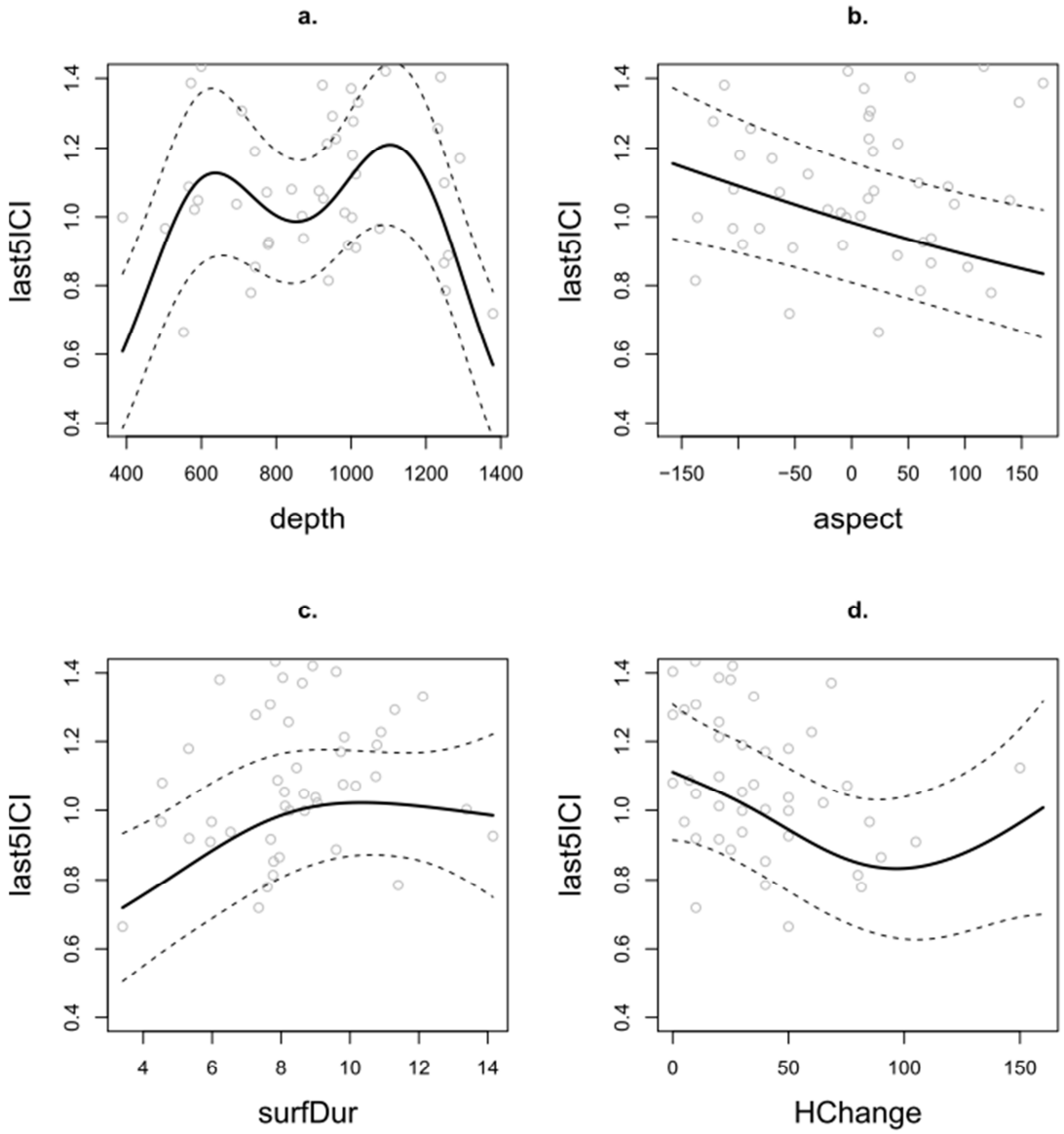


Figure 2.17 Predicted mean ICI (s) at the end of the first bout

Predictions are given as a function of fluking depth (a), sea bottom aspect at fluking (b), surface duration (c) and heading change (d), with no whale-watching vessels in the previous surfacing, individual fixed to its intercept and all other covariates fixed to their means (mean depth 885 m) under Model 15. Dashed lines: 95% confidence interval assuming t-distribution, grey circles: data.

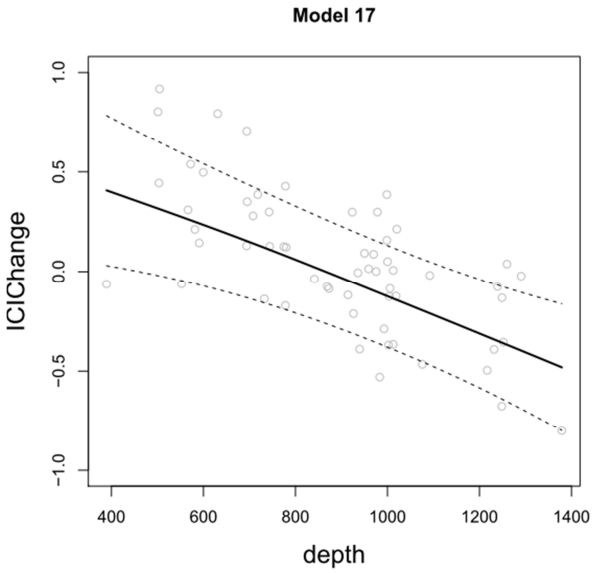


Figure 2.18 Predicted change in ICI during the first bout

Predictions are given as a function of fluking depth, with no whale-watching vessels, individual fixed to its intercept and all other covariates fixed to their means under ww-vessel Model 17. Dashed lines: 95% confidence interval assuming t-distribution, grey circles: data.

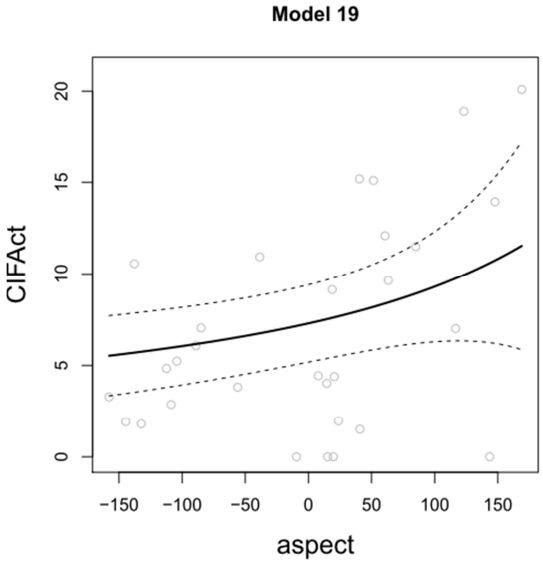


Figure 2.19 Predicted creak activity overlaid with data

Predictions are given as a function of sea bottom aspect, with no whale-watching vessels in the previous surfacing, individual fixed to its intercept and all other covariates fixed to their means under Model 19. Dashed lines: 95% confidence interval assuming t-distribution, grey circles: data.

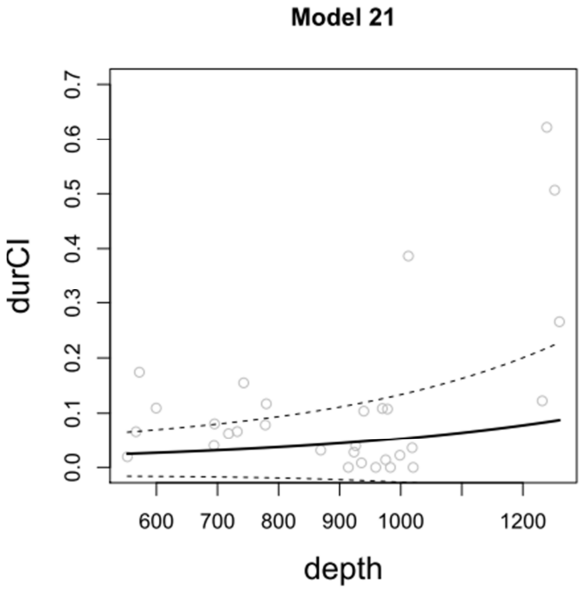


Figure 2.20 Predicted proportion of time spent creaking overlaid with data

Predictions are given a function of fluking depth, with no whale-watching vessels in the previous surfacing, individual fixed to its intercept and all other covariates fixed to their means under Model 21. *Dashed lines: 95% confidence interval assuming t-distribution, grey circles: data.*

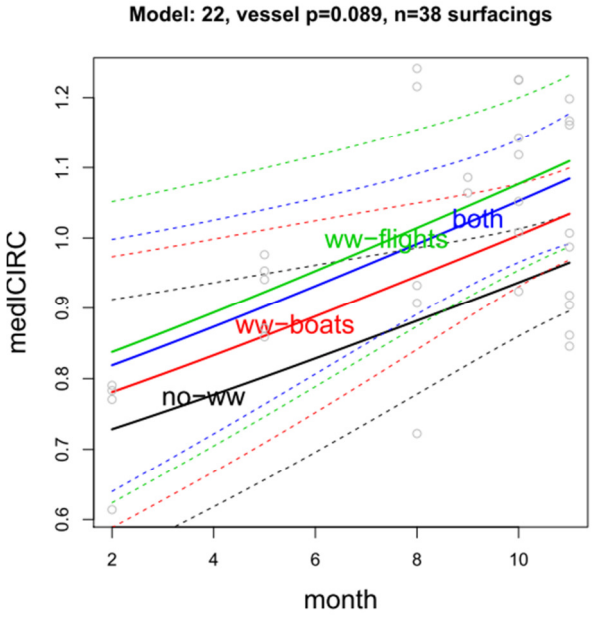


Figure 2.21 Predicted median ICI (s) after encounters with different whale-watching vessels

Predictions are given as a function of month, with mean values for all other covariates and individual factor fixed to its intercept under Model 22 (Table 8 a). Dashed lines: 95% confidence interval assuming t-distribution, grey circles: data.

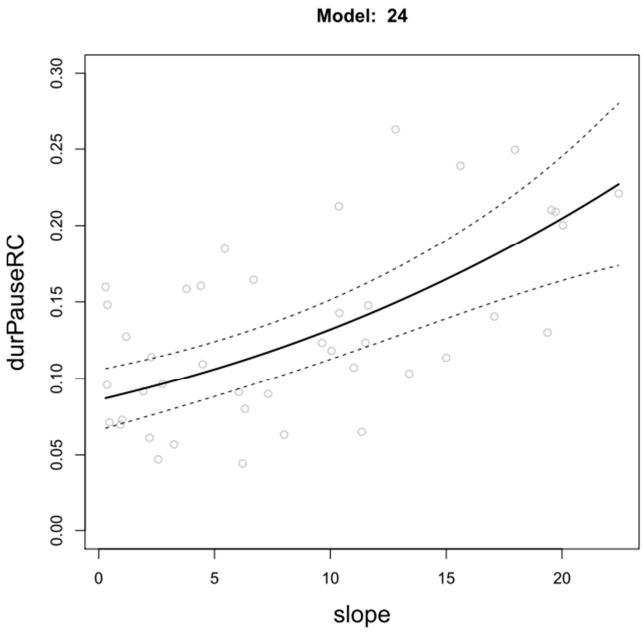


Figure 2.22 Predicted proportion of time spent silent

Predictions are given as a function of sea bottom slope at fluking, with no whale-watching vessels in the previous surfacing, individual fixed to its intercept and all other covariates fixed to their means under Model 24. Dashed lines: 95% confidence interval assuming t-distribution, grey circles: data.

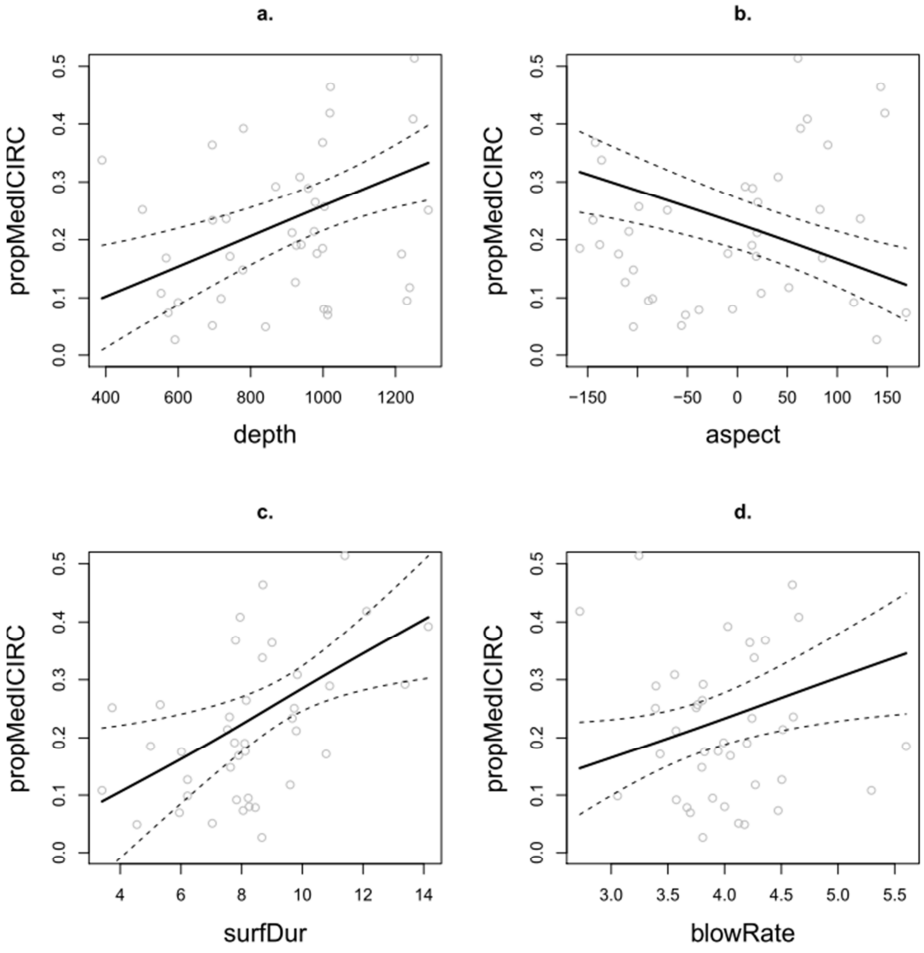


Figure 2.23 Predicted proportion of clicks in median ICI

Predictions are shown as a function fluking depth (a), sea bottom aspect (b), surface duration (c) and blow rate (d) given no whale-watching vessels in the previous surfacing, individual fixed to its intercept and all other covariates fixed to their means under Model 23. Dashed lines show 95% confidence interval assuming t-distribution. Grey circles show data.

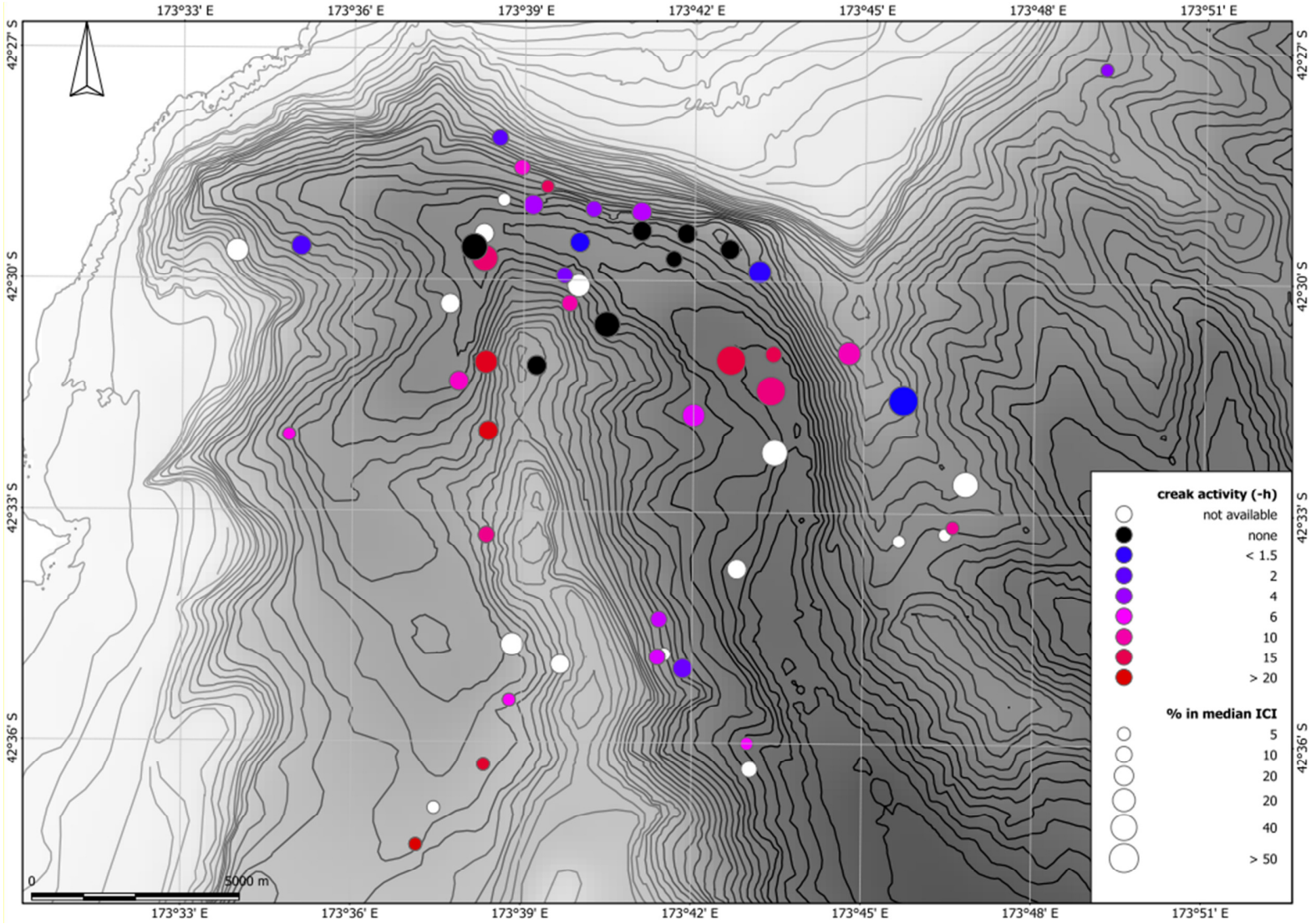


Figure 2.24 Distribution of data on creak activity (-h) and proportion of clicks near the median ICI

3 SPERM WHALE RESPONSE TO TAG BOAT PRESENCE: BIOLOGICALLY INFORMED HIDDEN STATE MODELS QUANTIFY LOST FEEDING OPPORTUNITIES

Abstract

Animal-attached sensors provide invaluable data to describe the natural behaviour of cryptic species, such as cetaceans, and are increasingly used to assess anthropogenic disturbance effects. Tag deployment and handling may themselves alter the behaviour of study animals and there is a need to assess if and when behaviour recovers close enough to an undisturbed level. Not all behavioural changes have fitness consequences (Chapter 2), and the goal here is to derive metrics that can be directly linked to fitness implications, such as time and energy allocation. In the present chapter, a hidden state-switching model is developed that incorporates biological knowledge and multiple streams of tag-recorded data to estimate time series of functional behavioural states for 12 sperm whales off Norway. Foraging, recovery and resting states were specified in the hidden state model by state-dependent likelihood structures. Comparison of hidden state models revealed a parsimonious set of input time series, and supported the inclusion of a less informed 'silent active' state. There was a high agreement between state estimates and expert classifications. The estimated functional time budget was then used in time series models (GLM & GEE) to test three hypotheses for behavioural change during remote suction-cup tag deployment procedures: change in behavioural states, and change in foraging success or locomotion cost, given behavioural state. Sperm whales spent less time at the sea surface (-34%) and more time in non-foraging silent active state (+60%) in the presence of the tag boat than during post-tagging baseline period. Time-decaying models of tagging effects were not retained in model selection, indicating a short-term effect that ceased quickly after the tag boat left the whale. No changes could be detected in energetic proxies, given behavioural state, but changes in functional state budget indicated costs in terms of lost feeding opportunities and recovery time at surface. These results are useful to quantitatively identify data periods that should not be considered baseline behaviour within tag recordings. This functional state approach proves an effective way to quantify disturbance in terms of time and energy allocation that is based upon general principles that can be applied to other species and biologging applications.

3.1 Introduction

Animal-attached sensors have become an important means to monitor individual behaviour for a wide range of species and habitats in the wild. With technological advances in miniaturisation, resolution and longevity of biologging sensors and transmitters, there is scope for a more integrated understanding of how individual behaviour and physiology interact with their environment and anthropogenic stressors (Cooke et al 2004, Hooker et al 2007, Johnson et al 2009). As such, biologging science can provide first clues of individual-level mechanisms that could drive anthropogenic impacts on populations (Cooke et al 2004, Hooker et al 2007, Tyack 2009, Berger-Tal et al 2011, Miller et al 2012). For population-level inferences to be reliably made, it is important to consider how representative the tagged individuals' baseline behaviour (such as time spent foraging) or response to stimuli (such as probability of avoidance) are of the wild, non-tagged population of conservation interest. Evaluation of possible effects of biologging experimental procedures is therefore important when considering how representative tag data might be to the entire population ("measurement affects performance", Wilson et al 1986, Miller et al 2009).

Research effects of biologging studies comprise both the effects elicited by the tag deployment procedures, such as approach, physical contact or capture (hereafter collectively termed as 'handling') and the presence of device itself on the animal (Murray & Fuller 2000, Baker & Johanos 2002, Hawkins 2004). Documented tagging and marking effects range from injury, physiological stress and behavioural changes to reproductive success and survival rates (Murray & Fuller 2000, Godfrey & Briant 2003, Barron et al 2010, Walker et al 2012). The relative importance of handling and device effects depends upon their relative invasiveness, duration and repetition that may allow for habituation or sensitization. The effects of tag presence are of particular concern for flying and swimming species that may be more sensitive to alterations to their streamlining, such as tag-induced drag (Bannasch et al 1994, Hazekamp et al 2009, Barron et al 2010), and subsequent increases in transport costs (Wilson et al 1986, Ropert-Coudert et al 2000, Wilson & McMahon 2006, Fossette et al 2008). These effects are reduced by use of relatively smaller and more aero- and hydrodynamic tag shapes (e.g.

Bannasch et al 1994). Locomotion costs can also be expected to increase if tag significantly increases the mechanical loading (weight), buoyancy or centre of gravity of an individual (Wilson et al 1986). Tag attachment method (e.g. harness vs. glue) may also impair movement (Barron et al 2010), but also have more subtle physiological effects, such as changes in the distribution of animal surface temperature (McCafferty et al 2007).

In marine mammals, most studies have reported short-term behavioural effects with little evidence of impacts on survival (Walker et al 2012, e.g. McMahon et al 2008). While extensive research on tagging effects have helped to guide deployment practices and tag development (e.g. Fossette et al 2008), generalising the device-specific and mostly qualitative results to different species and constantly evolving telemetry set ups is challenging (Murray & Fuller 2000). Not only are tagging effects likely to depend upon specific handling procedures and tag design but also individual (age, sex, condition) and behavioural and environmental context (e.g. nursing, prey availability) (Murray & Fuller 2000, Hawkins 2004, Walker et al 2012, Section 1.1). Reliable estimation of tagging effects therefore requires case-by-case assessment. However, with limited availability and cost of alternative study platforms, tagging studies are rarely able to empirically cross-validate tag data with data from a 'pre-tagging' period or data from non-tagged individuals (Murray & Fuller 2000, Godfrey & Briant 2003, Walker et al 2012). Most studies therefore assume that tagging has negligible or no influence on parameters of interest after some cut-off recovery time since handling (‘baseline’ period; Murray & Fuller 2000, Godfrey & Briant 2003, but see definition of baseline period based upon affected dive parameters in Miller et al. 2009).

An alternative and quantitative approach is to compare tagged individual behaviour between different available 'doses' of tagging procedures, such as varying tag size (Wilson et al 1986) or handling intensity (Engelhard et al 2002). Such an approach could be used to back-calculate true population parameters (Wilson et al 1986, Wilson and McMahon 2006). For example, Ropert-Coudert et al (2007) compared diving and movement behaviour of Adelie penguins between two different tag sizes to extrapolate effects on penguins with tags of negligible size. The authors predicted that non-tagged penguins would maintain similar energy

expenditure than tagged animals but be able to swim faster, dive deeper, and range farther in pursuit of prey. Similarly, data can be compared within each tag record under the assumption that handling effects are strongest at the time of attachment and decrease afterwards. For example, Miller et al (2009) found that the first dive after tagging was shorter than subsequent dives of sperm whales. Such a 'during-after' comparison can reduce confounding individual variability, but does assume that tag records are long enough to allow at least partial recovery.

In this chapter, a novel approach is developed to quantitatively assess the effects of suction-cup tag deployment procedures ('handling') on sperm whales for which no pre-tagging control was available. The objective was to compare whale behaviour in the presence vs. absence of the tag boat, and to evaluate different models of recovery from effects due to tag attachment and tag boat presence. Three classes of possible behavioural effects were evaluated: 1) change in behavioural state transitions, 2) reduction in foraging success, given behavioural state and 3) increase in movement cost, given behavioural state.

To obtain behavioural states for hypothesis testing, multiple streams of tag data were used in a hidden state-switching model to estimate biologically informed states and their uncertainty. As well as classification of sperm whale behaviour, the model development aimed to formulate 'functional' states that could be generalized to other species and used to assess a range of disturbance stimuli (Fig. 1.1). State-based approaches are promising to quantify behavioural context in behavioural response studies, such as responsiveness at the onset of exposure (Goldbogen et al 2013). More fundamentally, functional states can be used to capture the success-failure rate of harvesting a currency across different behavioural or motivational states. Time budgets can then be used to scale up any incurred costs of behavioural disturbance (e.g. Williams et al 2006), and define boundaries for carry-over effects (O'Connor et al 2014).

Sperm whales perform deep (200-1000m) and long (30-60min) foraging dives (Watwood et al 2006), facing trade-offs between time spent foraging at depth and recovering oxygen stores at the sea surface (Boyd 1997). These trade-offs formed the conceptual basis for the functional state model for sperm whales. Two bio-energetic currencies were considered, foraging success and movement cost, that vary across the foraging dive cycle (surfacing, descending transit,

layer-restricted search, ascending transit). Terminal echolocation buzzes (Miller et al 2004) and 'overall dynamic body acceleration' (ODBA; Halsey et al 2009) were quantified as proxies for foraging success and locomotion cost, respectively. Besides foraging dive cycles, sperm whales also spend time in shallower dives for other purposes, such as resting or 'silent active' swimming. Sperm whale resting dives occur in consecutive bouts of variable duration, are typically shallower than foraging dives, and are stereotypically characterised by a vertical 'head-up' or 'head-down' posture (Miller et al 2008). Non-foraging but active behaviours are also described for sperm whales (Miller et al 2008), and likely reflect social or anti-predatory functions (Curé et al 2013). The model was able to test how many non-foraging functional states are utilized by sperm whales by comparing models with 5 (foraging states + resting) versus 6 states (foraging states + resting + active non-foraging).

3.2 Materials and methods

Time series of functional states were estimated first, and the resulting state classification was then used to estimate behavioural disturbances likely linked to individual fitness. Behavioural states were estimated in a hidden state model in order to formalise prior expectations of functional behaviour (surfacing, transiting, layer-restricted search, resting, and other 'silent active') and utilise multiple input data time-series. The state estimates and uncertainty were then used as data in a second time-series model that tested for time or energetic costs of tag deployment procedures with different models of recovery from disturbed to baseline behaviour.

3.2.1 Data

Data were collected for 12 individual sperm whales tagged with an audio and movement-recording bio-logging device (DTAG; Johnson et al 2009). 4 whales were tagged in 2005 (Teloni et al 2008) and 8 whales were tagged in 2008-2010 (Miller et al 2012) near Lofoten Islands in Northern Norway. Sperm whales were localized at sea visually and acoustically by monitoring their echolocation clicks with a towed hydrophone array. The protocol included initial observations at 200-1000m from a main observation vessel (MS Stronstad, 29 m). A smaller tag

boat (rigid-hulled inflatable boat or similar) was launched to approach each whale and deploy tags with a poles that varied in length each year of research (Table 3.1).

Tag data were processed to calculate depth as well as whale-frame acceleration and magnetometer data which was converted to pitch, roll and heading time-series (Miller et al 2012). Time-series data from the tag was down-sampled to 1 sample per minute to reduce computational time and concentrate analysis efforts on dive phase scale rather than fine-scale behaviour, such as thrusting strokes. Depth was sampled at the start of each 1-minute interval, while mean pitch and 'overall dynamic body acceleration' (ODBA) were calculated over the entire 1-minute interval. ODBA was calculated as the sum over each minute of the 2-norm of high-pass filtered acceleration (finite impulse response filter, cut-on frequency 0.05 Hz). To normalize deployment effects on ODBA such as those due to variable tag position, ODBA values were divided by the whale's median value and then multiplied by the median ODBA across whales. Surface periods were detected using a depth threshold of 2 metres for accepting a dive, and a threshold of 1 m for reaching the surface. Time (min) since the last surface period was calculated for the start of each 1-minute interval ($minFromSurf$).

Audio data (stereo at 96 kHz) were monitored aurally and visually using spectrograms for echolocation click trains (regular and buzz clicks) and marked for their start time and duration in each record. The presence or absence of these aurally monitored clicks in each 1-minute interval was used in the hidden state models in conjunction with the depth and accelerometer data. The presence or absence of prey capture attempts within each 1-min interval was scored using the start time of buzzes. Other types of clicks (slow clicks, codas) were not included in the analysis.

Six whales were exposed to experimental naval sonar, silent pass by the transmission vessel as a control, and/or killer whale playbacks including noise controls (Miller et al 2012, Curé et al 2013). Two of these six whales were exposed to just one killer whale and one noise playback, followed by a secondary camera tag deployment 1.2 hours after all experiments ended. All data from all 12 individual sperm whales were used to parameterise the hidden state model, but non-tagging baseline periods excluded all exposures and post-exposures to sonar (experimental

or incidental), and killer whale playback experiments. For tagging effects analysis, tag handling periods were defined as the time period between tag deployment and recovery of tag boat to the main research vessel or movement of the tag boat far (>1km) from the tagged whale. Full experimental protocols are described in Chapter 4 (Section 4.2.2).

A calibration data set of behavioural states was used to compare with the hidden model state estimates. “Bottom phases” were defined by the period between the first positive and the last negative pitch in a dive for 9 whales (Miller et al 2004; Table 3.1). Dive types were classified by consensus of three experts, including the author, Dr Patrick Miller and Dr Stacy DeRuiter. The resulting consensus comprised 11 dive types (Table 3C.4).

3.2.2 Hidden state model

The state-switching model for sperm whale behaviour consisted of four functional foraging states and either one or two additional states for non-foraging related behaviour (Fig. 3.1). Alternative model structures were considered to assess how many states (five or six, Fig. 3.1) and which combinations of input data (depth, clicking, time since surfacing, ODBA and/or pitch) should be included to classify the behavioural time series most effectively. Each model consisted of a 5x5 or 6x6 state transition probability matrix and state-dependent likelihoods for the input data.

Depth at each time bin was modelled as a random walk Gaussian variable with a state-specific mean and variance (Photopoulous 2013, Langrock et al 2013):

$$d_t \sim N(d_{t-1} + \pi_{s_{t-1}}, \sigma^2{}_{s_{t-1}}) \qquad \textit{Equation 3.1}$$

where d_t denotes depth at time step t and s denotes the hidden state at time step $t-1$. Descent and ascent states were modelled as a directional random walk (‘bias’ parameter π_s estimated ≠ 0), and all other states a non-directional random walk ($\pi_s = 0$). A separate variance for depth changes ($\sigma^2{}_s$) was estimated for each state. A step function was used to constrain predicted depths to be > 0 m.

To relax the Markov assumption that state transitions depend only upon the previous time step, all models allowed the probability of surfacing at time t to increase with decreasing depth at time $t-1$ in a multinomial logistic regression (see Langrock et al 2013 for a similar formulation of feed-back in transition probabilities). $minFromSurf$ (x_1) was an additional covariate in the regression for the probability of transition to LRS (state 3). The linear predictor for the probability of state s at time t was therefore:

$$f(P(s_t)) = \beta_{0,s_{t-1},s_t} + \beta_{1,s_t} d_{t-1} + \beta_{2,s_t} * x_{1,t} \qquad \textit{Equation 3.2}$$

where intercept β_{0,s_{t-1},s_t} was specific to a state-transition, coefficient $\beta_{1,s_t} d_{t-1}$ was associated with transitions to surface (state 1), and β_{2,s_t} associated with staying in LRS (state 3). The coefficients were fixed at zero for other transitions, i.e. when $s_t \neq 1$, then β_{1,s_t} was set to zero, and when $s_{t-1} \neq 3$ and $s_t \neq 3$, then β_{2,s_t} was set to zero.

The presence/ absence of clicking (c) was estimated a state-specific probability ($c_t \sim \text{Bernoulli}(\gamma_{s_t})$). ODBA (o) was similarly modeled as a Gamma distributed variable with state-dependent shape and rate parameters ($o_t \sim \text{Gamma}(\varphi_{s_t}, \omega_{s_t})$).

The absolute value of the pitch angle p was modelled in a logistic Beta regression (Ferrari and Cribari-Neto 2004) so that within mobile states (i.e. not surfacing or resting), pitch was related to vertical step length in a linear predictor:

$$g(p_t) = \alpha_{0,s_t} + \alpha_{1,s_t} |d_{t-1} - d_t| \qquad \textit{Equation 3.3}$$

Here, the coefficient for vertical step α_{1,s_t} was specific to each state so that all mobile states were estimated a single coefficient which was fixed at 0 for surface and resting. Pitch during surfacing and resting were estimated state-dependent means ($\alpha_{0,1}, \alpha_{0,5}$), while mobile states were assigned a common intercept.

The joint likelihood for the full model (all five data streams) was the product of their conditionally independent likelihoods (for a similar formulation, see McClintock et al 2013):

$$l(\pi_s,\ \sigma^2, \beta, \gamma, \varphi, \omega, \alpha, \tau_s, s \mid \theta) = \qquad \textit{Equation 3.4}$$

$$\prod_{t=1}^{T} l(\pi_t|\theta, s_t)l(\sigma^2{}_t|\theta, s_t)l(\beta_t|\theta, s_t)l(\gamma_t|\theta, s_t)l(\varphi_t|\theta, s_t)l(\omega_t|\theta, s_t)l(\alpha_t|\theta, s_t)l(\tau_t|\theta, s_t)l(s_t|\theta, s_{t-1})$$

Where θ denotes the included set of state-dependent parameters. The full model (all four model components) had 54 estimable parameters in addition to the hidden states that were estimated for each data point.

After initial inspection of model performance, one additional parameter was introduced. In the fore-mentioned models, a time-constant average step length was assumed within each state by estimating a state-specific $\sigma^2{}_s$. Inspection of the data revealed that step lengths increased as a function of the depth during foraging dives (dives consisting of only descent, LRS, and ascent). The observed relationship appeared to be linear when depth was square root transformed (see Fig. 3.4, middle panel). A time-varying $\sigma^2{}_s$ was therefore specified for LRS state and time-varying drift for descent and ascent states by setting:

$$\sigma^2{}_{3,t} = \bar{\sigma}^2{}_3 + \mu * \sqrt{d_{t-1}} \qquad \textit{Equation 3.5}$$

$$\pi_{2,t} = \bar{\pi}_2 + \mu * \sqrt{d_{t-1}}$$

$$\pi_{4,t} = \bar{\pi}_4 + \mu * \sqrt{d_{t-1}}$$

Here $\bar{\sigma}^2{}_3$ and $\bar{\pi}_s$ are the time-constant intercepts for variance and drift for the random walk, $\sigma^2{}_{3,t}$ and $\pi_{s,t}$ the respective time-varying parameters, and μ the increase in step length for every square root unit increase in depth. The relationship between step length and depth was therefore specified constant across the three foraging states (descent, LRS, ascent). These models are hereafter referred to as time-varying step length (TS) models as opposed to the fixed step length (FS) models. For an exhaustive list of model parameters and script, see Appendix 3A.

In order to incorporate prior information on whale behaviour, a Bayesian approach was taken to parameterize the models. A Gibbs sampling algorithm was used to sample from the joint posterior distribution of the model. Freely available jags software (2003) was used within r (coda package, Plummer 2007 and R2jags package Su and Yajima 2012). Descent and ascent

rate were specified with informative priors using Gamma distribution with a mean and variance parameter from literature (Watwood et al 2006). A lower mean and variance for ODBA was used to construct a Gamma prior for resting. Probability of clicking was also informed, with a higher mean for foraging states (descent, LRS, ascent). Pitch regression coefficients had uninformative priors with no parameter difference between states except that the coefficient for vertical step was fixed at 0 for surface and resting as explained above. Uniform (uninformative) priors were specified for most transition probabilities (state-specific intercepts). Coefficients for the probability of transition to surface and LRS were assigned uninformative normal priors. The probability of transition to surface was constrained to be negative by truncating its prior distribution. See Appendix 3A for a comprehensive list (illustration in Fig. 3D.1).

All models were sampled in 3 independent chains, each with an initial 16,000 iterations. Model convergence was assessed at this point, and a subset of models that were deemed to reach convergence in terms of state classification were updated a further 20,000 times. Initial values were set manually for all state parameters (Table 3A.1). Brooks-Gelman-Rubin diagnostic (BGR; Brooks & Gelman 1998, Gelman et al 2003) was used to assess model convergence, which was rejected based on its poorest converging parameter (BGR estimate ≤ 1.05). Detailed methods and results for model convergence can be found in Appendix 3B.

Four criteria were used to rank models that were deemed to have converged: 1) goodness of fit relative to model complexity (deviance information criterion DIC), 2) uncertainty in state classification, and 3) sensitivity and specificity to pre-classified bottom phases and 4) comparison to pre-classified resting and silent-active dives.

Deviance information criterion (DIC) was used to measure goodness of fit relative to model complexity. The DIC is an extension of Aikake's Information Criterion (AIC) and is particularly useful for models that have been fitted outside of a maximum likelihood (ML) framework. Similarly to the AIC, the DIC is based upon both model fit and model complexity (Spiegelhalter et al 2002, Lunn et al 2013):

$$DIC = \overline{D} + p_V \qquad \textit{Equation 3.6}$$

where $\overline{D}$ is the posterior mean deviance of the model, and p_D is the effective number of model parameters. Jags calculates the deviance as the sum of the deviances of all observed random variables defined in the model (i.e. "stochastic nodes" in BUGS terminology), and p_D as the difference between the expected deviance $\overline{D}$ and the deviance evaluated at the posterior means ($D(\bar{\theta})$; Spiegelhalter et al 2002). However, p_D cannot be evaluated for discrete parameters such as hidden states (Lunn et al 2013). An alternative measure of effective number of parameters p_V was used instead, which is invariant to reparameterisation but assumes that the information in the likelihood dominates that of the prior (Gelman et al 2003, Lunn et al 2013):

$$p_V = {}^{\mathrm{var(D)}}/_{2} \qquad \textit{Equation 3.7}$$

An assessment of the goodness of fit was made between the posterior state-dependent likelihoods and data, or loosely termed "certainty" in state classification. The joint probability density of data ("emission probability") and probability of state transitions were calculated for each model given the posterior parameter samples. The emission and transition probabilities were calculated for each data point as per model specification, but ignoring prior distributions. Transition probability matrix was updated at each time step to incorporate the linear predictor with data on depth and time since surfacing (minFromSurf). The "emission only" prediction was calculated by selecting the state that maximised the sum of the log-likelihoods for data (i.e. in the full model, the likelihood for depth, clicking, pitch and ODBA). The "Viterbi sequence" accounted for both emission and transition probabilities by calculating the likelihood for the entire sequence using the Viterbi algorithm (Viterbi 1967, see Appendix 3A script). The predicted states were then compared to the posterior state estimates to assess the contribution of state-dependent likelihoods vs. transition probabilities in the state classification. The two estimates are expected to differ because the posterior state-dependent likelihoods (data) should not always support the expected states based on the sequence of

states, e.g. a data point resembling drifting (state 5) in the middle of an estimated layer-restricted search phase.

To measure the contribution of data in a given state classification, the most likely states were re-calculated based on a sub-set of emission probabilities from the full model. The full model was chosen in order to compare the contributions of all data streams. The predicted states based on emission probabilities were compared to the model's state estimates, and the % of correct predictions for each state was contrasted across the sub-sets.

Layer-restricted search (LRS) state estimates were compared to pre-classified bottom phases and drifting state and silent active state estimates to expert classification of dives (Table 3C.4) to assess their concordance to existing methods of behavioural classification. Unlike LRS state, bottom phases were limited to a single phase within each dive that started and ended with changes in descend and ascend pitch. Therefore a higher sensitivity of LRS state to bottom phases could have also indicated a classification that was less sensitive to multi-layered dives.

Measures of accuracy (Metz 1978) and diagnostic odds ratio (DOR, Glas et al 2003) were used to compare LRS state to pre-classified bottom phases. The sensitivity of LRS state estimates to pre-classified bottom phases was calculated as the total proportion of LRS state estimates within bottom phases. The specificity of LRS state estimates was calculated as the proportion of non-LRS state estimates within the whole time-series that was not classified as outside bottom phases.

DOR combines sensitivity and specificity into one discriminatory test performance diagnostic (Glas et al 2003):

$$DOR = \frac{\frac{sensitivity}{1-sensitivity}}{\frac{1-specificity}{specificity}} \qquad Equation\ 3.8$$

Thus, DOR was the ratio of the odds of LRS state in a bottom phases to the odds of LRS state outside the bottom phase. The higher the value, the better the estimated state could discriminate between the human classified states. Standard errors were calculated for the

logarithm of DOR that follows approximately a normal distribution (Glas et al 2003). Sensitivity and specificity of the LRS state classification to bottom phases were calculated based upon the proportion of bottom phases ('true conditions') and intervals between bottom phases ('false conditions') that were estimated as LRS state. The number of bottom phases were therefore accounted for in $SE(\log DOR)$.

3.2.3 Use of state classifications for assessment of tagging effects

The top-ranked time series estimate of hidden states was used as data, and their uncertainty as weights, in a second analysis step that tested the effects of tagging on three response variables: 1) $state$ - estimated activity state (~ multinomial), proxy for functional state, 2) $buzz$ - presence/absence of buzzing (~ Bernoulli), proxy for foraging success, and 3) $ODBA$ - overall dynamic body acceleration (~ Gamma), proxy for locomotion cost.

Probability of $state$, given previous state, was modelled by including previous state ($prevState$) as factor baseline covariate. $State$ was used as a factor baseline covariate in models for $ODBA$ and $buzz$. Mean differences were allowed in all three response variables across individuals by including tag id ($whale$) of the record as a factor covariate. The binomial model for $buzz$ was fitted to a subset of data that only included foraging states (descent, LRS and ascent). No buzzing was observed in the non-foraging states (surfacing, drifting or silent active), so estimating standard errors for their coefficients would have not been possible.

Candidate exposure covariates for tagging were assessed for inclusion using model selection, and were designed to test between different hypotheses of the time-course of possible behavioural responses to tagging (Fig. 3.2). Presence/absence of tag boat was included either as a main effect ($Tagging$), or interaction with year ($Tagging : year$) or pole length ($Tagging : poleL$) to assess any differences in level of response across years or as a function of pole length, respectively.

A maximum likelihood framework was chosen for fitting these models for ease of model selection using AIC. Multinomial log-linear regression models were fit using function multinom in r library nnet, while binomial (logit link) and Gaussian (identity link) regression models were

fit using function glm in r library stats. Multinomial models were weighted by the posterior probability of the state estimate, thus accounting for the uncertainty in state estimation. AIC unit difference of ΔAIC < -2 was considered support for candidate tagging covariates compared to the baseline models for each response variable ($state \sim prevState + whale$, $ODBA \sim state + whale$, and $buzz \sim state + whale$). All tagging effects models included the baseline covariates and up to two explanatory variables for Tagging. To avoid spurious relationships, only one of the four time-decay covariates ($minFromTd$, $minFromTd2$, $minFromTagging$, or $minFromTagging2$) were included in any one model, and were not included in the same model with $Tagging:Year$.

The lowest AIC models were diagnosed for influential individuals and data, goodness of fit, distributional assumptions, and serial correlation in residuals (Appendix 3E). Models that were diagnosed with serial correlation of residuals were re-fit within a generalised estimating equation (GEE) in SAS 9.3 (procedure 'genmod'). In multinomial models, the state that appeared to change most in response to tagging was used as a binomial response variable in the GEE. Any tagging effects were re-assessed using the empirical standard error estimates that do not assume any particular working correlation within the GEE, but account for the smaller effective sample size of correlated data within clusters (Gosho 2014). Small empirical standard errors (estimates > 2 x SE) and significant type 3-tests ($p<0.05$) were considered as support for candidate covariates. GEE models included $whale$ as a cluster variable rather than an explanatory variable, and therefore explicitly estimated the parameters of the model for the group of whales rather than separately for each individual.

3.3 Results

3.3.1 Data

A total of 175.37 hours of DTAG data were analysed, an average of 14.6 hours of data recorded per whale (Table 3.1).

3.3.2 Hidden state model

MODEL CONVERGENCE

Based on their state classification convergence at 6k-16k iterations, 8 fixed-step length (FS) models were rejected and 10 accepted for further updates. All 6-state FS models that did not include pitch failed to converge in terms of state classification, suggesting that pitch was important in discriminating between resting (state 5) and active-silent state (state 6). In the 10 FS models selected to be updated, all parameters converged adequately (BGR estimate < 1.05) after 16,000 iterations (Table 3B.2).

The 7 lowest DIC model structures (Table 3.2) were also fitted with time-varying step length (TS) during foraging states descent, LRS and ascent. Of the 7 TS models, 2 models with 5 states (models 2 and 6) failed to converge in terms of state classification; the remaining 5 TS models and their parameters appeared to converge sufficiently (Table 3B.3). The converged set of models improved within-chain correlation of all posterior transition probabilities from state 3 to states 2-6 compared to the same models without TS (Fig. 3B.3). Detailed description of model convergence can be found in Appendix 3B.

DEVIANCE INFORMATION CRITERIA

6-state models had both lower posterior mean deviance and DIC than their respective 5-state model structures. Conversely, posterior mean deviances were higher for models with ODBA despite increased model complexity (Table 3.2). The effective number of parameters was estimated small for models with smaller deviance (model structures 6 and 9) and higher for models with higher deviance (model structures 4, 5 and 8) both by p_D and p_V (Table 3.2, Fig. 3C.1). Therefore, deviance and DIC arrived at a similar ranking of models.

STATE CERTAINTY

Time series of state estimates were calculated for each model as the most prevalent state in the posterior sample at iterations 16-36k for FS models and 24-48k for TS models. The state estimates from all the 15 models agreed on 77.9 % of data, yielding similar time budgets (Table 3C.1, Table 3C.2). 6-state model classifications were more consistent within fixed step length

(FS) models (93.9%) and within time-varying step length (TS) models (93.9%) than across (86.7%). TS models estimated the highest proportions of data as LRS state than any FS model (>40%, Table 3C.2).

'Overall state uncertainty' was designed to measure the average residual or overall lack of support for the estimated sequence of states. Overall state uncertainty ranged between 3.3-4.5% of samples across all models. Allowing for step length to increase with depth improved the overall state certainty in all converged model structures (Fig. 3C.2). TS models 5 (full model) and 8 (pitch + ODBA) had the lowest overall state uncertainty (3.25% and 3.33%).

Based on emission probability of data alone (depth, clicking, pitch, ODBA), the models' discriminatory power broadly mirrored that of their overall state uncertainty (Fig. 3C.2, Fig. 3C.3), indicating that any lack of support for the most prevalent states was driven by the state-dependent likelihoods. Emission probabilities predicted a posterior average of 89.1% to 93.45% of state estimates across models. When accounting for transition probabilities (Viterbi algorithm), the models' ability to discriminate states was improved and less variable between models (97.8% to 98.9%). Viterbi algorithm improved the state predictions only 7.0% on average, highlighting how variable and relatively little the (mostly) Markov state-transitions contributed to the state classification (Fig. 3C.3).

Surface and drifting states had the lowest average state uncertainties (0.44% and 0.70%) and descent and ascent states the highest across all models (6.08% and 5.48%) (Table 3C.3, Fig. 3C.4). Silent active state had similarly high average uncertainty both within FS and TS models (5.74 and 6.07%, Table 3C.3). Although the emission probabilities predicted silent active state better than the foraging states 2-4 (Fig. 3C.4), silent active state was predicted worse than any other state when accounting for transition probabilities (Fig. 3C.5). However, excluding the 6th state from the model appeared to decrease the contribution of state-dependent likelihoods in descent state estimation (Fig. 3C.4). The state-dependent likelihoods for TS models were further better able to discriminate descent and ascent states than FS models (Fig. 3C.4). The overall lower state uncertainty of TS models therefore appeared to be driven by the foraging states (descent, LRS and ascent).

Contribution of data was measured for the state classification of the full TS model structure 5 with 6 states. Compared to the full set of likelihoods, the % of correct predictions decreased most for LRS state and silent active state when clicking was excluded from the predictions. In contrast, removing ODBA changed the % of correct predictions least (Fig. 3C.6).

Comparison to pre-classified resting and silent-active dives

There were only small differences in the estimated time budgets for expert classified (Table 3C.4) drifting dives between the models. All models estimated drifting dives to consist more than an average of 76% (76.3-82.9%) of time in drifting state, and at least an average of 13% (12.9-18.1%) of time in state 2 (descending). 5-state models with pitch estimated drifting dives to also contain ascending (3.9-4.4.6%) while 6-state models estimated silent active (silent active state, 7.0-7.9%) (Fig. 3C.7, Table 3C.5).

Expert classified silent active swimming dives had more variable time budgets across models than drifting dives. 5-state models without pitch estimated these dives to consist mostly of drifting state (Model 2 average: 75.2%, Model 4 average: 68.7 %) while 5-state models with pitch estimated these dives to consist mostly of ascend (state 4 averages 61.2-62.6%) (Fig. 3C.7, Table 3C.5). 6-state classifications were more consistent, with ~75% in silent active state and ~3% in drifting state.

Sensitivity and specificity to pre-classified bottom phases

LRS state estimates of model structure 6 and 9 with 6 states were most sensitive to the pre-classified bottom phases (0.79 and 0.78 respectively, Fig. 3C.8, Table 3C.5). With little differences in specificity between the models, also the diagnostic odds ratio (DOR) selected for these two models as the best match for pre-classified bottom phases. In terms of the 95% confidence intervals for DOR, all FS models appeared to be significantly poorer classifiers of bottom phases than TS models (Fig. 3.3).

Selection of a final model

6-state and TS models outperformed respective 5-state and FS models, both in terms of lower DIC, lower state uncertainty and higher sensitivity to pre-classified bottom phases (Fig.

3.3). 6-state models estimated most of the time in expert classified 'silent active swimming' dives as silent active state, and models with pitch were further able to discriminate between expert classified drifting and silent active swimming dives (Fig. 3.3). When vertical step was allowed to vary with depth (TS models), inclusion of ODBA appeared to somewhat improve overall state certainty and sensitivity to pre-classified bottom phases (Fig. 3.3).

Minimum DIC was obtained for model structures 6 ($base + pitch$) and 9 ($base + pitch + minFromSurf$) both within 5- and 6-state models. However, both uncertainty in state classification and sensitivity to pre-classified bottom phases ranked three models slightly above the lowest DIC model (6-states, TS and $pitch$): full 6-state TS model, and 6-state TS models $pitch + minFromSurf$ and $pitch + ODBA$ (Fig. 3.3). Including $ODBA$ in the best DIC model with pitch changed only 2.8% of its state estimates, a magnitude similar to their overall state uncertainty (~3%), and had only small contribution on the state classifications of the full TS model (Table 3C.2, Fig. 3C.6). In the interests of model parsimony therefore, ODBA was selected against in the hidden state model. Including $minFromSurf$ in the best DIC model changed the state classification even less, by 0.6%. Without $minFromSurf$, TS model posterior samples, transition probabilities from state 3 in particular, had a greater (>>400) effective sample size. Therefore, it was the lowest DIC model 6 ($base + pitch$) with 6 states and time-varying step length that was selected for interpretation and further analyses of tagging effects.

3.3.3 Description of selected model

The posterior distributions of the selected hidden state model were consistent with prior expectation of behaviour. A high probability of clicking was estimated for the foraging states (posterior means for descent: 0.90, layer-restricted search: 0.99, and ascent: 0.56) while a low probability of clicking was estimated for surface, resting and silent active states (<0.02). Descent and ascent rates overall were very similar when accounting for their variability and effects of depth (Fig. 3.4). During foraging states (descent, LRS and ascent), step length was estimated to increase by 1.47 (SD 0.02) m/min for every unit increase in square root transformed depth. The posterior mean absolute value of pitch was 1.3 (SD 45.6) degrees

during surfacing and 80.5 (SD 45.8) degrees during resting. See Appendix 3D for complete description of the selected model.

3.3.4 Effects of tagging

Data for tagging effects analysis included 9 DTAG deployments (87.62 h) from the time of first tag-on to the first experiment or end of the full tag record (Table 3.1). For two of these whales, the period between the start of secondary tag deployment until the end of the full DTAG record was also included. Three whales (sw05_199a-c) were excluded completely due to incidental exposures to unidentified sonar at the beginning of the tag records (0-3 hours from tag deployment).

When the tag boat remained near the whales in tagging operations (n=8.1 h), the whales spent no time resting, and across individuals, an average of 1.6x more time in the silent active state (10.1%, SD=13.2) and less time surfacing (12.4%, SD=10.2) compared to baseline periods when the boat was recovered (n=79.4 h; 60% increase from 6.3%, SD=15.1 and 34% decrease from 18.8%, SD=5.5 respectively) (Fig. 3.6).

The most prolonged tagging period was for whale sw08_152a that was approached by the tag boat for 2.8 hours after tag attachment attempting to photograph the whale (Fig. 3.5). During those 2.8 hours, the whale spent only 1.8% of the time in surfacing state 1, and 31.8% of the time in silent-active state. With most of the silent-active state comprised silent diving, the whale spent only 12.3 % of its time near surface (< 10 metres). Immediately after the tag boat left the whale, it spent 8 minutes in the surfacing state, which was the longest period the whale spent in the surfacing state during the entire DTAG record.

The lowest AIC model for state transitions included $prevState + whale + Tagging$, which improved the baseline model $prevState + whale$ by 9.5 AIC units (Fig. 3.7). $Tagging$ covariate was also supported by a likelihood ratio test between the two models (df=5, p=0.002, function anova.nnet()). The model estimated 86.5% of the post $Tagging$ baseline states and 79.2% of the $Tagging$ period states correctly. The model fit best to LRS and drifting states (92.7% and 88.5% correct predictions, respectively), and worst to silent active state (64.8%) (Fig. 3E.1).

The binomial GEE model for silent active state with *prevState* and *Tagging* as covariates and *whale* as a cluster variable improved the QIC of the baseline GEE model by 28.9 units (Table 3.3). The GEE model with *Tagging* estimated the odds of silent active state to increase by a factor of 3.70 [95% CI 1.3, 10.1] during *Tagging* (Table 3E.4), slightly greater but within the confidence intervals (2xSE) of the respective coefficient estimate from the multinomial model (Table 3E.3). These GEE model results confirmed that the detected change in state transitions of the multinomial model was not merely a by-product of serial correlation. Both positive and negative residual correlation was detected in the best multinomial model (Fig. 3E.2).

Probability of buzzing was highly variable within and across individuals, but the individual average for foraging states was somewhat lower during *Tagging* (descent state: 8.4% SD 14.4; LRS state: 15.0%, SD= 11.1; ascent state: 0.0%) than in the post *Tagging* baseline (descent state: 10.1 % SD 8.6; LRS state: 23.1% SD 14.6; ascent state: 5.8% SD 4.5) (Fig. 3.6). There was no consistent increase in *ODBA* during *Tagging* compared to baseline across states. Only surface and ascent states had slightly greater individual average *ODBA* during *Tagging* (surface state: 26.7 SD 7.4; ascent state: 22.0 SD 5.5) than post *Tagging* (surface state: 21.9 SD 3.5; ascent state: 20.6, SD 3.5) (Fig. 3.6).

In the AIC model selection, there was little support for an overall tagging effect on probability of buzzing, given the foraging states (descent, LRS and ascent states) and *whale* as a factor covariate. *Tagging* improved the baseline model *state* + *whale* by only 0.74 AIC units (Fig. 3.7). The coefficient estimate for *Tagging* was small (-0.32, SE 0.20) with no evidence that it was different from zero (z=-1.61, p=0.107).

The best AIC model for ODBA included *state* + *whale* + *minFromTd*. *minFromTd* improved the baseline model by 30.0 AIC units (Fig. 3.7), however, the estimated effect was very small (-0.18 decrease in mean *ODBA* for every hour). When fitted within a GEE which accounts for serial correlation in the data, neither *minFromTd*, *Tagging* nor *Tagging:state* were supported with respective QIC increases of 543.7, 18.5 and 1353.22 units compared to the

base model (Table 3.3). Therefore, there was little evidence for a change in $ODBA$ as a function of time since tag-on time or tag boat presence.

3.4 Discussion

In this chapter, a time-series of functional behavioural states was estimated that most fully captured the variability in diverse data streams recorded by an animal-attached movement and sound recording tag. These results demonstrate that a 'silent active' state can be identified despite lack of prior information, and that including a silent-active state along with a defined resting state improved the functional state models for behaviour of Norwegian sperm whales. The output of the model was then used to demonstrate that whales spent more time in the non-foraging 'silent active' state when the tag boat was present, and that a simple present versus absent response explained the data better than time-decaying models of behavioural response. This enables quantitative determination of post-handling periods that should be excluded to retain periods more likely to reflect baseline behaviour (used to contrast behaviour during experiments in Chapter 4).

3.4.1 Hidden state models

The hidden state models were able to estimate both very stereotyped states (surfacing, resting) and states with highly variable data signatures (layer-restricted search, other non-foraging) (Fig. 3.4). Although there was uncertainty in formal model selection in this Bayesian framework, different hidden state models arrived at similar state classifications, which all agreed well with expert classifications (Fig. 3.3). The hidden state models succeeded in identifying and characterising states that could be interpreted in terms of functional behaviours previously documented for sperm whale foraging. The accepted hidden state model included 6 states, time-varying vertical step length for foraging states (descent, layer-restricted search and ascent), clicking, and log-linear relationships between vertical step and the absolute value of pitch. This parsimonious model had the lowest DIC score, and almost the lowest uncertainty and agreement with expert opinion as other models (Fig. 3.3).

Allowing step length to increase with depth (TS models) improved the DIC, state uncertainty, and sensitivity to pre-classified bottom phases compared to models with a simple random walk with fixed step length (FS models) (Fig. 3.3). TS models also appeared to capture an active foraging mode better than FS models that estimated the highest average ODBA during descent rather than LRS. These results do not suggest that sperm whales are intrinsically more mobile at depth, but rather that the time-varying formulation for step length was more flexible by accepting a wider distribution of step lengths for LRS, and was subsequently able to more fully capture an active foraging mode. Such high variability in step length across foraging phases could be expected when prey layers vary in vertical thickness, quality and/or prey species that in turn influence whales' hunting and searching strategy (Fais et al in press 2014).

3.4.2 Functional time budget of foraging male sperm whales

Layer-restricted search (LRS) was estimated as the most prevalent state in the post-tagging data (47.5% of all data, and 51.2% of all foraging states 1-4), consistent with the high proportion of time spent in foraging and high diving efficiencies (foraging phase duration: dive cycle duration) reported for sperm whales both at high- and low latitudes (Jaquet et al 2000, Watwood et al 2006). Unlike studies using bottom phase (defined by the first descent and final ascent of a dive) or foraging phase (defined by the first and last buzz of a dive) (Watwood et al 2006) alone as a measure of foraging time however, the hidden state model allowed multiple foraging phases could be estimated within a dive (Appendix 3D, e.g. Fig. 3D.3 b). 30 out of 119 (25.2%) "usual" foraging dives (expert dive types 1-4 in Table 3C.4) contained more than one foraging (LRS) phase.

There was strong support for a sixth 'silent active' state, with 6-state models outperforming 5-state models in terms of higher overall posterior probability of states (Fig. 3.3) and a better fit of state-dependent likelihoods to the data (Fig. 3C.3). Furthermore, there was high concordance between the state 6 estimates and expert classified "silent active" dives (Fig. 3.3). Nevertheless, state transitions appeared to be relatively weak predictors of state 6 compared to other states (Fig. 3C.5), with wide posterior credible intervals for the transition probability of

staying in state 6 (Fig. 3D.1 a). However, variability related to state transitions is expected as state 6 likely encompassed several non-foraging behaviours that may have been associated with different functional behaviours and contexts, such as socialising, avoiding the tag boat near surface, or horizontal transit. Future work with larger datasets could consider the potential to divide state 6 into more specific functional states.

State 5 (resting/ drifting) was estimated for 3.8% of post tagging baseline data, most of which coincided with expert classified drifting dives based upon the description of this behaviour by Miller et al (2008). For two whales (sw05_196a and sw10_150a) state 5 also identified drifting to the sea surface that occurred at the end of foraging dives (max depth 306 metres, Fig. 3D.3 a). Drifting had a very distinct data signature featuring little vertical movement yet nearly vertical pitch (posterior mean and 95% CRI was estimated for step length as 8.5 [7.9, 9.0] m, and for pitch as 80.5 [79.8, 81.0] degrees), consistent with stereotyped vertical posture drift-dives documented for sperm whales world-wide (Miller et al 2008).

3.4.3 Effects of tagging

Using the estimated states and uncertainty to assess tagging effects, sperm whales were found to increase time in non-foraging silent active state (Table 3.3, Figs. 3.6, 3.7). Within each behavioural state, there was little evidence of changes in a proxy for locomotion cost (ODBA) or a proxy for foraging success (probability of buzzing) (Table 3.3, Figs. 3.6, 3.7). These results indicate a direct evasion or vigilance reaction to the tag boat that disrupted behaviour, rather than reduced foraging success or locomotion cost when whales did enter foraging states in the presence of the tag boat. No longer-term effects could be detected on the time scale of each tag record (~15-20 h in duration), suggesting that whales recovered to a post-tagging level of behaviour almost immediately (within minutes) after the tag boat left.

Comparable data could not be collected during the pre-tagging period, and therefore it could not be established with certainty that behaviour was resumed to a completely undisturbed (non-tagged) level. However, two whales (sw10_149a and sw10_150a) were re-approached for a secondary tag deployment and had a response profile consistent to whales that were tagged

only once. Both whales spent time in the silent active state near the sea surface during first and second Tagging periods, while full foraging dive cycles were resumed soon after the tag boat left the whales. Within both tag records, the first and second post-tagging deployment periods consisted of near identical time and depth profile of foraging (descent, LRS and ascent states) with little apparent effect of the presence of a secondary tag (Fig. 3D.3 k-l). It was therefore concluded that the presence of tags alone on the animal was likely to have little effect on whale behaviour compared to tag deployment procedures ('handling'). Indeed the DTAG only weighs 300 g, which is less than 0.01% of an adult sperm whale mass (14-50 t, Best 1979). Little is known about the effects of suction-cup tag attachment, however tags typically detach if a sperm whale performs a breach, indicating that an uncomfortable tag can be removed by the whale (Johnson et al, 2009).

Although there was little evidence for any changes in energetic proxies within states, the increased probability of non-foraging silent active behaviour and reduced time at surface suggests an energetic cost of tag boat presence. Miller et al (2009) found similar short-term changes in sperm whale foraging behaviour during the first dive of the tag record but not subsequent dives. These changes included reduced buzz and pitching rates during the bottom phase, and shorter dive duration compared to the subsequent dive. However, the presence/absence of the tag boat was a more important predictor of effects than time since tag-on time, suggesting a lack of a specific cut-off period after tag attachment. This result is expected when the tagging procedure, including re-approach for photo-identification, varies across tagging occasions. In such cases it is important to collect detailed data on tagging effort to describe the 'dose' of handling, such as tag boat distance to the whale, with focus on recording the intensity and duration of approach both before and after a successful tag attachment.

3.4.4 Methods considerations

Although there was little evidence of short-term tag boat effects, the sensitivity of the test for subtle longer-term effects was likely to be limited due to the relatively small number of tags

(n=9) and high variability in state budgets and buzz rates across the tag records. Variability in the tagging procedure across years and different tagging crews was also likely to affect the probability and level of the individual responses. No evidence was found for a different level of response to shorter pole length (Fig. 3.7), however due to the small sample size, not all factors could be accounted for that could have been equally important, such as tag boat handling and targeting tag placement near the head vs. the back of the animal.

Tagging periods after tag attachment ranged from just 6 minutes up to 2.8 hours (1.1 to 61.4% of tag records). The time-series approach explicitly modelled this unbalanced sample, and results were also contrasted from GLMs that estimated individual level differences with GEEs that estimated individual average and between-individual variability in the response data. Nevertheless, it was possible that a few individuals that responded strongly to the presence of tag boat were influential in the estimation of a population effect. The influence of individuals was tested by re-fitting the baseline and the tagging effects GLM:s for state without each individual, and found that excluding either sw08_152 or sw10_150a lowered the AIC difference below the ΔAIC threshold of -2 (Fig. 3E.4). sw08_152 was exposed to the longest Tagging period (2.8 h), whereas sw10_150a was approached by tag boat twice for shorter periods, including secondary tag deployment. Both sw08_152 and sw10_150a spent the longest durations in state 6 (an average of 6.5 and 4.0 minutes, respectively) compared to any other whale during Tagging, and were not estimated to return to state 6 in the baseline period. Therefore, had these two apparently more sensitive individual not been sampled, a tag boat effect would have gone undetected.

As one of the first attempts at multivariate hidden state modelling of individual behaviour, the hidden state model structure was simplified by assuming mostly Markov transitions, no individual effects or spatial memory for prey layers. Despite the relatively simple process model, a sufficiently strong signal in the input data allowed for robust state classification and estimation of time budgets that were highly variable across individuals. A more realistic (complex) process model would be required if disturbance was incorporated and tested within the hidden state model. For example, a hidden state model with individual as a random effect

could estimate population-average effects by incorporating tag boat presence as an explanatory variable for buzzing within each state.

3.4.5 Conclusions

The functional state approach appears to be able to effectively estimate behavioural disturbances that can be linked to individual fitness. These results showed that after tag deployment, whales can remain vigilant to the presence of tag boat and thus trade off foraging time for perceived risk at surface. During-after comparisons of functional states and currency proxies were influenced by individuals that were exposed to tag boat repeatedly or for extended durations, highlighting the importance of consistent deployment procedures and minimizing handling time. Nevertheless, these data showed a cut-off point (tag boat recovery) after which whales were likely to have returned to a post-tagging level of behaviour. These results lend support for the exclusion of handling periods to better define post-tagging baseline data that can be considered as a more accurate sample of undisturbed behaviour. However, in order to estimate how representative the baseline behaviour of tagged individuals is of the non-tagged population of interest, comparable pre-tagging data are needed. An optimal design would monitor behaviour before, during and after tagging from a remote platform that minimised research effects, and collect behavioural data that can be complemented by onboard tag record. For cetaceans such as sperm whales that use biosonar to locate prey, visual and passive acoustic tracking would be a promising avenue to monitor foraging and movement at a coarse scale that could be calibrated by the onboard acoustic and orientation record.

3. References

Baker JD, Johanos TC (2002) Effects of research handling on the endangered Hawaiian monk seal. Mar Mammal Sci 18:500-512

Bannasch R, Wilson RP, Culik B (1994) Hydrodynamic aspects of design and attachment of a back-mounted device in penguins. J Exp Biol 194:83–96

Barron DG, Brawn JD, Weatherhead PJ (2010) Meta-analysis of transmitter effects on avian behaviour and ecology. Methods Ecol Evol 1:180–187

Beale CM (2007) The behavioral ecology of disturbance responses. Int J Comp Psychol 20:111–120

Best PB (1979) Social Organization in Sperm Whales, Physeter macrocephalus. In: Behaviour of Marine Animals (eds. Winn HE, Olla BL). Plenum press, New York, US

Berger-Tal O, Polak T, Oron A, Lubin Y, Kotler BP, Saltz D (2011) Integrating animal behavior and conservation biology: a conceptual framework. Behav Ecol 22:236–239

Boyd I (1997) The behavioural and physiological ecology of diving. Trends Ecol Evol 12:213-217

Brooks S, Gelman A (1998) General methods for monitoring convergence of iterative simulations. J Comput Graph Stat 7:434–455

Cooke SJ, Hinch SG, Wikelski M, Andrews RD, Kuchel LJ, Wolcott TG, Butler PJ (2004) Biotelemetry: a mechanistic approach to ecology. Trends Ecol Evol 19:334–43

Curé C, Antunes R, Alves A (2013) Responses of male sperm whales (Physeter macrocephalus) to killer whale sounds: implications for anti-predator strategies. Sci Rep 3:1579

Engelhard GH, Hall AJ, Brasseur SMJM, Reijnders PJH (2002) Blood chemistry in southern elephant seal mothers and pups during lactation reveals no effect of handling. Comp Biochem Physiol A Mol Integr Physiol 133:367–78

Fais A, Soto NA, Johnson MP, Pérez-González C, Miller PJO, Madsen PT (In press, 2014) Sperm whale echolocation behaviour reveals a directed, prior-based search strategy informed by prey distribution

Ferrari S, Cribari-Neto F (2004) Beta Regression for Modelling Rates and Proportions. J Appl Stat 31:799–815

Fossette S, Corbel H, Gaspar P (2008) An alternative technique for the long-term satellite tracking of leatherback turtles. Endanger species Res 4:33-41

Gelman A, Carlin JB, Stern HS, Rubin DB (2003) Bayesian data analysis. Chapman and Hall/CRC, Boca Raton, Florida, US

Gill J, Norris K, Sutherland W (2001) Why behavioural responses may not reflect the population consequences of human disturbance. Biol Conserv 97:265–268

Glas AS, Lijmer JG, Prins MH, Bonsel GJ, Bossuyt PMM (2003) The diagnostic odds ratio: a single indicator of test performance. J Clin Epidemiol 56:1129–1135

Godfrey J, Bryant D (2003) Effects of radio transmitters: Review of recent radio-tracking studies. Sci Conserv 214:83–95

Gosho, M (2014) Criteria to select a working correlation structure for the generalised estimating equations method in SAS. J Stat Softw 57:1-10

Halsey LG, Shepard ELC, Quintana F, Gomez Laich a, Green J a, Wilson RP (2009) The relationship between oxygen consumption and body acceleration in a range of species. Comp Biochem Physiol A Mol Integr Physiol 152:197–202

Hawkins P (2004) Bio-logging and animal welfare: practical refinements. Mem Natl Inst Polar Res 58:58–68

Hazekamp AAH, Mayer R, Osinga N (2009) Flow simulation along a seal: the impact of an external device. Eur J Wildl Res 56:131–140

Hooker SK, Biuw M, McConnell BJ, Miller PJO, Sparling CE (2007) Bio-logging science: Logging and relaying physical and biological data using animal-attached tags. Deep Sea Res Part II Top Stud Oceanogr 54:177–182

Jaquet N, Dawson S, Slooten E (2000) Seasonal distribution and diving behaviour of male sperm whales off Kaikoura: foraging implications. Can J Zool 1990:407–419

Johnson M, Aguilar de Soto N, Madsen PT (2009) Studying the behaviour and sensory ecology of marine mammals using acoustic recording tags: a review. Mar Ecol Prog Ser 395:55–73

Langrock R, Marques T, Baird R, Thomas L (2013) Modeling the diving behavior of whales : a latent-variable approach with feedback and semi-Markovian components. J Agric Biol Environ Stat:1–19

Lunn D, Jackson C, Best N, Thomas A, Spiegelhalter D (2013) The BUGS Book. Chapman and Hall/CRC, Boca Raton, Florida, US

McCafferty DJ, Currie J, Sparling CE (2007) The effect of instrument attachment on the surface temperature of juvenile grey seals (Halichoerus grypus) as measured by infrared thermography. Deep Sea Res Part II Top Stud Oceanogr 54:424–436

McClintock B, Russell D, Matthiopoulos J, King R (2013) Combining individual animal movement and ancillary biotelemetry data to investigate population-level activity budgets. Ecology 94:838–849

McMahon CR, Field IC, Bradshaw CJA, White GC, Hindell MA (2008) Tracking and data–logging devices attached to elephant seals do not affect individual mass gain or survival. J Exp Mar Bio Ecol 360:71–77

Miller PJO, Aoki K, Rendell LE, Amano M (2008) Stereotypical resting behavior of the sperm whale. Curr Biol 18:R21–3

Miller PJO, Johnson MP, Madsen PT, Biassoni N, Quero M, Tyack PL (2009) Using at-sea experiments to study the effects of airguns on the foraging behavior of sperm whales in the Gulf of Mexico. Deep Sea Res Part I Oceanogr Res Pap 56:1168–1181

Miller PJO, Johnson MP, Tyack PL (2004) Sperm whale behaviour indicates the use of echolocation click buzzes “creaks” in prey capture. Proc R Soc London Ser B Biol Sci 271:2239–47

Miller PJ, Kvadsheim PH, Lam, Frans-Peter A Wensveen PJ, Antunes R, Alves AC, Visser F, Kleivane L, Tyack PL, Sivle LD (2012) The Severity of Behavioral Changes Observed During Experimental Exposures of Killer (Orcinus orca), Long-Finned Pilot (Globicephala melas), and Sperm (Physeter macrocephalus) Whales to Naval Sonar. Aquat Mamm 38:362–401

Murray DL, Fuller MR (2000) A critical review of the effects of marking on the biology of vertebrates. In: Boitani L, Fuller T.K. (eds) Research Techniques in Animal Ecology. Columbia University Press, New York, p 15–64

Nathan R, Getz W (2008) A movement ecology paradigm for unifying organismal movement research. Proc Natl Acad Sci 105:19052–19059

O’Connor C, Norris D, Crossin G, Cooke S (2014) Biological carryover effects: linking common concepts and mechanisms in ecology and evolution. Ecosphere 5:1–11

Oliveira CC, Wahlberg M, Johnson M, Miller PJO, Madsen PT (2013) The function of male sperm whale slow clicks in a high latitude habitat: communication, echolocation, or prey debilitation? J Acoust Soc Am 133:3135–44

Photopoulous T (2012) Diving and depth use in seals: inferences from telemetry data using regression and random walk movement models. University of St Andrews

Plummer M (2003) JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. Proc 3rd Int Work Distrib Stat Comput (DSC 2003)

Plummer M, Best N, Cowles K, Vines K (2007) The CODA package, version 0.16-1, plummer@iarc.fr

Ropert-Coudert Y, Bost CA, Handrich Y, Bevan RM, Butler PJ, Woakes a J, Maho Y Le (2000) Impact of externally attached loggers on the diving behaviour of the king penguin. Physiol Biochem Zool 73:438–44

Ropert-Coudert Y, Knott N, Chiaradia A, Kato A (2007) How do different data logger sizes and attachment positions affect the diving behaviour of little penguins? Deep Sea Res Part II Top Stud Oceanogr 54:415–423

Spiegelhalter DJ, Best NG, Carlin BP, Linde A van der (2002) Bayesian measures of model complexity and fit. J R Stat Soc Ser B (Statistical Methodol) 64:583–639

Su YS, Yajima M (2012) R2jags: A Package for Running 'JAGS' from R. URL http://cran.r-project.org/web/packages/R2jags/R2jags.pdf.

Teloni V, Mark JP, Patrick MJO, Peter MT, Johnson MP, Miller PJO, Madsen, Madsen PT (2008) Shallow food for deep divers: Dynamic foraging behavior of male sperm whales in a high latitude habitat. J Exp Mar Bio Ecol 354:119–131

Tyack PL (2009) Acoustic playback experiments to study behavioral responses of free-ranging marine animals to anthropogenic sound. Mar Ecol Prog Ser 395:187–200

Viterbi A (1967) Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. Inf Theory, IEEE Trans 13:260–269

Walker KA, Trites AW, Haulena M, Weary DM (2012) A review of the effects of different marking and tagging techniques on marine mammals. Wildl Res 39:15

Watwood SL, Miller PJ, Johnson M, Madsen PT, Tyack PL (2006) Deep-diving foraging behaviour of sperm whales (Physeter macrocephalus). J Anim Ecol 75:814–825

Williams R, Lusseau D, Hammond PS (2006) Estimating relative energetic costs of human disturbance to killer whales (Orcinus orca). Biol Conserv 133:301–311

Wilson RP, Grant WS, Duffy DC (1986) Recording devices on free-ranging marine animals: does measurement affect foraging performance? Ecology 67:1091–1093

Wilson RP, McMahon CR (2006) Measuring devices on wild animals: what constitutes acceptable practice? Front Ecol Environ 4:147–154

3. Tables & Figures

Table 3.1 Summary of tag records

tag id		Sample duration (h)			% at surface	% at bottom	% in expert dives			
	Pole (m)	total	tagging analysis	tag-boat			D1-D4	D5-D7	D8-D9	D10-11
sw05_196a	15	21.32	21.32	0.50	29.7	52.0	91.6	5.8	0.0	2.7
sw05_199a	15	18.07	0.00	0.00	18.6	57.8	100.0	0.0	0.0	0.0
sw05_199b	15	13.82	0.00	0.00	22.9	46.1	82.9	6.6	10.5	0.0
sw05_199c	15	13.38	0.00	0.00	24.3	18.1	55.9	6.7	34.9	2.4
sw08_152a	5	8.65	4.60	2.83	16.2	no data	70.0	24.2	0.0	5.8
sw09_141a	9	15.28	3.83	0.82	20.7	no data	42.4	28.4	7.6	21.6
sw09_142a	9	14.77	2.98	0.23	21.0	no data	59.8	13.9	13.1	13.2
sw09_153a	9	8.53	8.53	0.12	17.6	61.7	100.0	0.0	0.0	0.0
sw09_160a	9	14.78	3.47	0.22	17.4	no data	94.7	2.9	0.0	2.4
sw10_147a	12	15.77	15.77	0.93	30.7	27.7	71.8	3.6	24.0	0.6
sw10_149a	12	16.13	14.15	1.80	21.9	51.8	95.9	0.0	0.0	4.1
sw10_150a	12	14.87	12.97	0.78	25.9	30.5	93.2	4.5	0.0	2.3
Total		175.37	87.62	8.23	266.8	345.6	958.3	96.6	90.1	55.0

Total sample duration (h) refers to data that was used to fit hidden state models, while tagging analysis show durations of data retained for tagging and post-tagging datasets (see text). Tag boat shows the total number of hours that the boat remained near the whale after tag deployment. Duration of pre-detected behavioural states (surface, bottom, and dive types) is shown as percentage of total sample size in each record.

Table 3.2 DIC selection criteria for hidden state models

Model #	state #	structure	Deviance	p_V	DIC
6 TS	6	depth + clicking + pitch	42654.9	4369.5	47024.4
9 TS	6	depth + clicking + minFromSurf + pitch	42767	4665.1	47432.1
9 FS	6	depth + clicking + minFromSurf + pitch	43478.9	4593.8	48072.7
6 FS	6	depth + clicking + pitch	43446.7	5122.8	48569.5
9 TS	5	depth + clicking + minFromSurf + pitch	46541.5	3393.7	49935.1
6 FS	5	depth + clicking + pitch	47463.5	3654.5	51118.1
9 FS	5	depth + clicking + minFromSurf + pitch	47505.7	3852.6	51358.3
2 FS	5	depth + clicking	74128.5	5614.6	79743.2
8 TS	6	depth + clicking + ODBA + pitch	118344.3	4658.2	123002.5
5 TS	6	depth + clicking + minFromSurf + ODBA + pitch	118495.1	4543.1	123038.1
5 FS	6	depth + clicking + minFromSurf + ODBA + pitch	119157.2	6587.2	125744.4
8 FS	6	depth + clicking + ODBA + pitch	119091	6682.1	125773.1
5 FS	5	depth + clicking + minFromSurf + ODBA + pitch	123024.8	3913.2	126938
8 FS	5	depth + clicking + ODBA + pitch	122975.6	4254.8	127230.4
4 FS	5	depth + clicking + minFromSurf + ODBA	151908.6	5713.6	157622.3

Deviance, effective number of parameters (p_V) and deviance information criterion (DIC) are shown for the 15 converged models in the last 10 000 iterations. Model numbers identify different model structures, 'FS' fixed step length models and 'TS' time-varying step length models.

Table 3.3 Model selection for Tagging effects

Response	Explanatory variables	Random effect	AIC / QIC	Δ AIC / Δ QIC
state	prevState + whale	-	4549.84	0.00
state	prevState + whale + Tagging	-	4540.38	-9.45
state 6	prevState	whale	1020.14	0.00
state 6	prevState + Tagging	whale	991.21	-28.92
buzz	state + whale	-	3136.38	0.00
buzz	state + whale + Tagging	-	3135.63	-0.74
buzz	state	whale	3253.02	0.00
buzz	state + Tagging	whale	3246.63	-6.39
ODBA	state + whale	-	38754.62	0
ODBA	state + whale + minFromTd	-	38724.62	-30.00
ODBA	state + whale + Tagging	-	38754.39	-0.23
ODBA	state	whale	191513.45	0
ODBA	state + minFromTd	whale	192057.18	+543.73
ODBA	state + Tagging	whale	191531.97	+18.52

The lowest AIC models (GLM; no random effects) are shown along with the corresponding GEE models with QIC (with whale as a random effect).

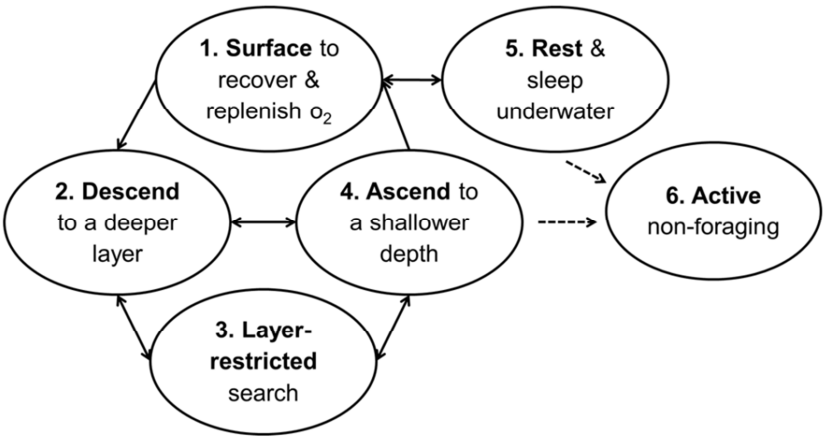


Figure 3.1. Concept model for sperm whale foraging behaviour

Five or six functional states were specified for sperm whales in their foraging ground: 1) surfacing, oxygen replenishment and physiological recovery at the surface; 2) descending transit, transiting to a deeper prey layer; 3) layer restricted search (LRS), searching at a prey layer; 4) ascending transit, transiting to a shallower depth or the surface; 5) resting and sleep underwater and 6) active non-foraging, which could encompass multiple functions. States 1-4 are considered to be functional states for foraging. Solid arrows show transitions that were expected to be likely and dashed arrows highlight the uncertainty related to the transition probability to and from state 6. These expectations and uncertainties were incorporated in the model as respective informative and uniform priors for the transition probabilities (Fig. 3A.1).

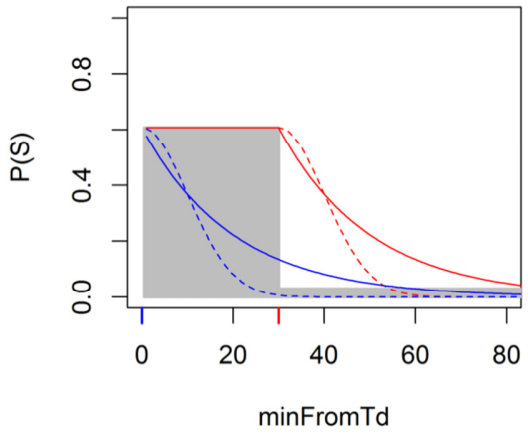


Figure 3.2. Different hypotheses for the 'dose' of Tagging

Functions illustrate the log-linear model probability of state transition $log(P(S)) = \alpha + \beta * x$ with five different hypotheses for tagging dose. Blue and red tick marks on x-axis show Tagging period, with start of tagging data in blue and end of $Tagging$ in red. The first hypothesis for dose was a presence/absence effect of tag boat, $Tagging$, shown as shaded grey. Four time-decaying explanatory variables were tested for hypotheses of recovery from either tag deployment (blue; $minFromTd$ and $minFromTd2$) or end of Tagging period (red; $minFromTagging$ and $minFromTagging^2$). The variables were calculated as linear or squared time since tag deployment or Tagging, representing either exponential (dashed lines $f(x)$; $minFromTd$ and $minFromTagging$) or exponential with delayed (dashed lines $f(x^2)$; $minFromTd2$ and $minFromTagging2$) speed of recovery. In this illustration example, the intercept α was set at -0.5 and coefficient β at -0.005.

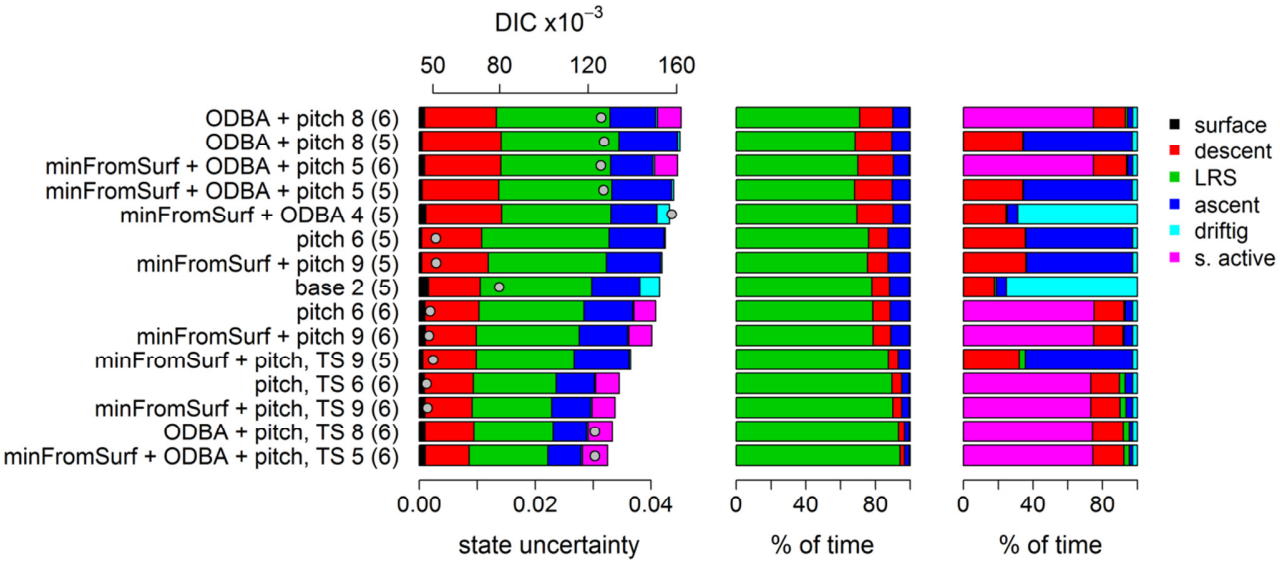


Figure 3.3 Hidden state model selection

Model structure numbers are given before number of states in brackets. 'Base' structure here refers to *depth + clicking* that were included in all of the converged set of models; 'TS' refers to time-varying step length models. Left panel: overall state uncertainty for each model (total bar width) with contributing states colour-coded. Overall state uncertainty was calculated for each model as the total proportion of posterior samples that were not the most prevalent state. Grey circles show DIC (from Table 2). Middle panel: % of time estimated in each state during pre-classified bottom phases. Contributing states are colour-coded so that green shows sensitivity of layer-restricted search to pre-classified bottom phases. Right panel: % of time estimated in each state during expert classified silent active dives. Models in all panels are shown in ascending order for overall state uncertainty.

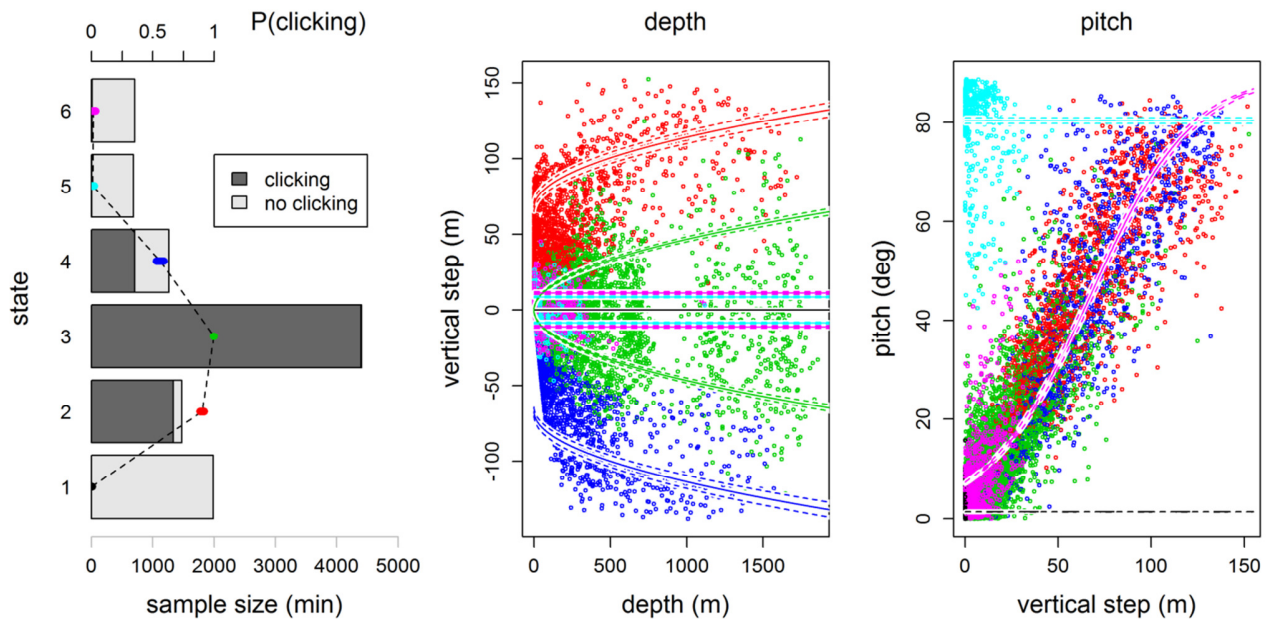


Figure 3.4. Characteristics of the selected functional state model

Left panel: Sample size and posterior 95% quantile for probability of clicking by functional state (Fig. 1). The total numbers of states with and without clicking are given on the bottom grey x-axis, and the posterior estimate for the probability of clicking on the top black x-axis. Middle panel: vertical steps (m/min) predicted as a function of depth (m). Posterior mean steps as a function of depth were predicted based on the posterior mean (solid lines) and 95% quantile (dashed lines) for the random walk parameters ${\sigma^2}_s$, μ, π_2 and π_4 (Table 3A.1). Predictions for each state are colour-coded; vertical step predictions for descents (red) and ascents (blue) include drift (bias, π) and are slightly asymmetric around zero because descent and ascent drift were estimated separately as π_2 and π_4 in the model. Vertical step predictions for states 1 (surfacing, in black), 3 (LRS, in green), 5 (resting, in indigo) and 6 (active-silent, in pink) did not include drift (i.e. not signed) but for illustration, are overlaid here symmetrically with observations both above and below zero. Right panel: Absolute value of pitch (deg) predicted as a function of vertical step length (m/min) (right), each overlaid with observed data. Pitch values were predicted based on the posterior mean (solid lines) and 95% quantile (dashed lines) values the pitch regression intercept $\alpha_{0,s}$ and coefficient for depth α_{1,s_t} (Table 3A.1).

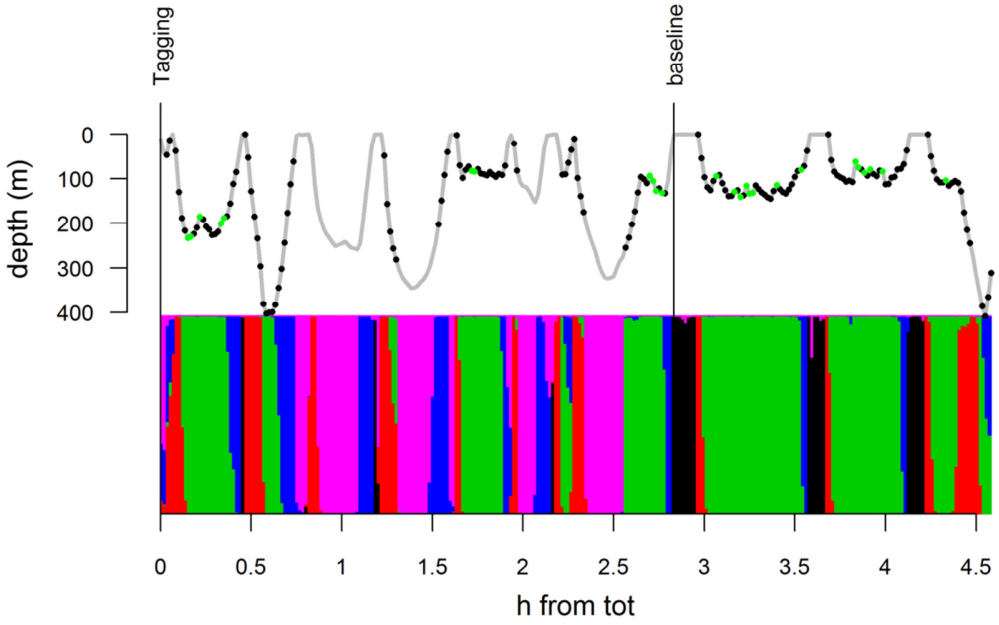


Figure 3.5 Example time series of states

Time series of state budget and dive profile for whale sw08_152 during the Tagging and post tagging baseline period. X-axis shows time since tag-on time (tot). Bottom graph shows posterior probabilities for each state. Top graph shows 1-minute depth data (grey) overlaid with presence/absence of clicking (black) and presence/absence of buzzing (green).

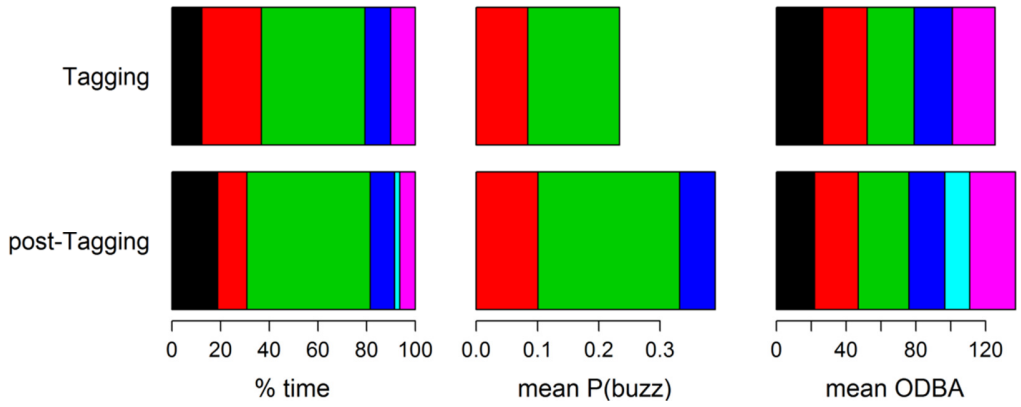


Figure 3.6 Time, buzz and ODBA budgets

Behavioural time budgets (left) and proxies of foraging success (probability of buzzing, centre) and locomotion cost (mean ODBA, right) averaged across individuals during the Tagging condition (top, 8.1 h) and post Tagging baseline (bottom, 79.4 h).

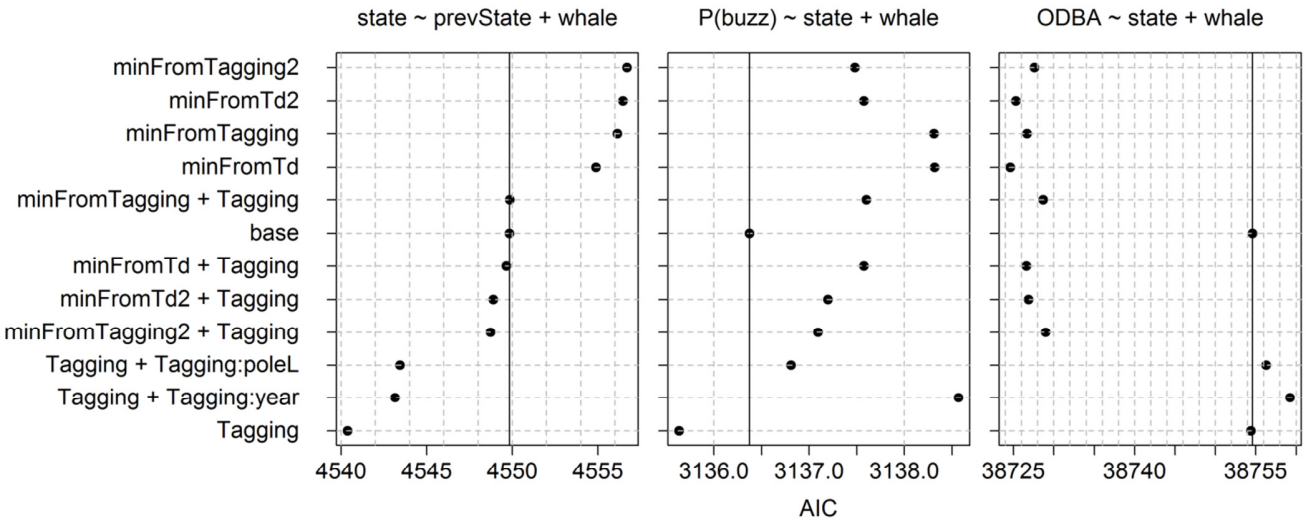


Figure 3.7 Model selection for Tagging effects

AIC model selection for tagging effects on state transitions (left) probability of buzzing (middle) and ODBA (right). Baseline models are shown on top of each figure. and candidate Tagging covariate combinations on the left. Black solid circles show AIC for each model and vertical line AIC value for the baseline model. Models were considered to have performed better than the baseline model if their AIC was at least two units lower (horizontal grid length).

4 SPERM WHALES REDUCE FORAGING EFFORT DURING EXPOSURE TO BOTH 1-2 KHZ NAVAL SONAR AND KILLER WHALE SOUND EXPOSURES

Abstract

The time and energetic costs of behavioural responses to incidental and experimental sonar exposures as well as playbacks of killer whale sounds were quantified using the estimated time-series of states and analytical framework developed in Chapter 3. Behavioural state transition modelling showed that tagged whales switched to a non-foraging 'silent active' state during both experimental transmissions of low-frequency active sonar (LFAS; 1-2 kHz, source level 214 dB re 1µPa m) from an approaching vessel (4 tag records) and playbacks of mammal-eating killer whale sounds from a stationary boat (5 tag records). Individuals switched to the silent active state at wide range of received sound pressure levels (SPL = 131-165 dB re 1µPa). Time spent in foraging states and subsequently the rate of echolocation buzzes was reduced during these two types of exposures with little change in overall locomotion cost, suggesting an effect on energy intake rather than on expenditure. In contrast, no changes in foraging behaviour were detected during experimental controls (silent vessel approach or stationary noise playback) or experimental medium frequency active sonar exposures (MFAS; 6-7 kHz, source level 199 re 1µPa m, received SPL = 73-158 dB re 1µPa). Similarly, there was no reduction in foraging effort for three whales exposed to incidental, unidentified 4.7-5.1 kHz sonar signals that were received at lower levels (SPL = 89-133 dB re 1µPa). These results appear inconsistent with a simple dose-response function with received SPL, and indicate that increased perception of risk with higher source level or lower frequency may modulate how sperm whales respond to anthropogenic noise.

4.1 Background

Cetaceans currently face rapid changes in their marine habitat such as noise pollution, exploitation of their prey species and offshore energy installations (Reeves et al 2003). Cetaceans rely upon sound for feeding, communication and navigation and are therefore thought to be especially vulnerable to anthropogenic noise (e.g. marine traffic, ships' sonar, seismic exploration) (Southall et al 2007, Tyack 2009). Effects of naval sonar are of particular concern due to high source levels, global geographic scale of exercises and associated stranding events of several species of beaked whales (Frantzis 1998, Filadelfo et al 2009), with other species also implicated (e.g. Balcomb and Claridge 2001) (D'Amico et al 2009). Research is focussed on the effects of active sonars that are typically hull-mounted or towed on anti-submarine warfare vessels (D'Amico et al 2009). Despite the fact that naval sonar frequencies (typically <14 kHz) are below the best hearing sensitivity and vocalisation ranges of most odontocete species (Southall et al 2007), research has found behavioural responses at a wide frequency range (eg. 1-2 kHz and 6-7 kHz, Miller et al 2012). Behavioural changes have been reported (Tyack et al 2011, DeRuiter et al 2013 a) at much lower levels of active sonar than would have been expected by the temporary hearing threshold shift (TTS) criterion for behavioural disruption (Southall et al 2007). These results indicate that factors other than direct injury may play a role in eliciting response to sonar at low received levels.

Hypotheses for causal mechanisms between sonar exposure and strandings range from the purely behavioural to those involving physiological stress and injury. Sonar may initiate a sequence of events offshore that could involve panicked flight or passive drifting towards shore, and/or perception of shallower water as a refuge. Deep-diving cetaceans have been suggested to be more susceptible to sonar due to physiological constraints related to diving metabolism and tissue recovery that may be compromised by a stress response and/or behavioural modifications such as a flight (Zimmer and Tyack 2007, Fahlman et al 2014). Cetaceans might not be fully immune to dysbaric osteonecrosis and decompression sickness (Moore and Early 2004, Fahlman et al 2014), and necropsies of sonar-related strandings suggest in-vivo formation of gas bubbles in vital organs (Fernández et al 2005). Direct acoustically-

mediated bubble growth has also been suggested (Fernández et al 2005) but in many stranding cases the received acoustic exposure appears to have been too low for such an effect (Fahlman et al 2014). The most extreme nitrogen tensions have been predicted for deep-diving beaked whales and sperm whales, in slow adipose tissues during long dives in particular, but shallow-diving species such as killer whales are also expected to have high blood and tissue supersaturation (Zimmer and Tyack 2007, Hooker et al 2009, de Quirós et al 2012, Kvadsheim et al 2012, Fahlman et al 2014). Risk of gas and fat embolism varies with a species-specific dive schedule, and the risk may be increased during sonar by dives that increase the time spent at depths where nitrogen is absorbed (Kvadsheim et al 2012). More energetic responses, such as the strong and sustained avoidance responses documented for Cuvier's beaked whales (DeRuiter et al 2013 a), may also compromise tissue recovery from anaerobic metabolism and initiate emboli by accumulation of carbon dioxide (Fahlman et al 2014). Therefore, energy expenditure may incur physiological costs, exhaustion and injury at much shorter time scales than what would be expected for terrestrial or shallow diving species.

Documented behavioural responses of cetaceans to experimental and operational sonar range from the presumably benign, e.g. blue whale or sperm whale orientation responses during 3.5-4 kHz active sonar exposures (Miller et al 2012, Goldbogen et al 2013) and vocal matching (false killer whales and pilot whale response to MFA, DeRuiter et al 2013 b, Alves et al 2014), to likely severe behavioural effects such as avoidance and cessation of foraging in killer whales (MFA, Miller et al 2012, 2014), pilot whales (LFA, Antunes et al 2014), beaked whales (MFA, Tyack et al 2011, DeRuiter et al 2013a), blue whales (Goldbogen et al 2013) and sperm whales (LFA, Miller et al 2012). Most of these studies have aimed to detect biologically relevant behavioural changes and to quantify exposure levels related to those changes. Recent advances, such as expert scoring of severity of behavioural responses (Miller et al 2012), predictive models for physiologically dangerous diving responses (Fahlman et al 2014), and multivariate behavioural break-point statistics (DeRuiter et al 2013 a, Miller et al 2014, Antunes et al 2014) have been critical to define a wide array of potential physiological and behavioural effects. However, the potential biological consequences of cetacean responses as fitness trade-

offs to perceived risk and the potential impact of more chronic or cumulative exposures of naval sonar remain poorly understood (Southall et al 2008, Clark et al 2009, Wright et al 2011).

This chapter sets out to estimate any functional time and energetic costs of sperm whale behavioural responses to naval sonar. Tagged sperm whales were subject to controlled exposure experiments (CEE) with both low frequency active sonar (LFAS 1-2 kHz) and medium frequency active sonar (MFAS 6-7 kHz) transmissions from an approaching source, vessel approaches without sonar transmission as a negative control, and mammal-eating *Orcinus orca* (killer whale) playbacks (Miller et al 2012). From this CEE dataset, behavioural responses to MFAS and LFAS have already been scored as minor to moderate by a consensus panel of experts (Miller et al 2012), and behavioural responses to playback of killer whale sounds have been established and interpreted as anti-predatory (Curé al 2012). The objective of the present study is to use quantitative classification of sperm whale behaviour to describe the relative time and energetic costs of behavioural changes during sonar exposures, and contrast any changes with presumed predatory context as a biologically relevant reference (Section 1.1; Frid and Dill 2002). In order to critically evaluate the baseline variability in sperm whale energetic budgets, an additional six tagged whales that were tagged in the area but not exposed to CEE or killer whale playbacks were included (total 12 DTAG:s, as in Chapter 3). Potential behavioural changes were also investigated for three of these whales that were exposed to previously unpublished incidental sonar (4.75-5.10 kHz).

Chapter 3 described a novel method to estimate functional behavioural states using acoustic, depth and orientation tag data in a hidden state switching model for sperm whale diving behaviour. In the present chapter, the estimated states and energetic proxies are used to quantify relative time and energy trade-offs across the experimental (CEE) and incidental sonar and control stimuli (killer whale playbacks and silent vessel approach). Based upon the estimated short-term effects of tagging procedures in Chapter 3, only post-tagging data were used as the pre-exposure baseline data against which to measure change during experimental sessions. The ’relative cost’ was defined as any reduction in foraging time or prey encounter

rate, or increase in locomotion effort, associated with the exposure relative to the pre-exposure (but post-tagging) baseline and control periods.

4.2 Methods

4.2.1 Data

Data from 12 sperm whales tagged in Norwegian waters that were processed in Chapter 3 were used in this analysis (Table 3.1). These included the time series of posterior distribution for hidden states (Chapter 3, Fig. 3D.3), presence/absence of a buzz start, and overall dynamic body acceleration (ODBA) at 1-minute intervals (Section 3.2.1). For detection of effects, data were excluded for all periods when the tag boat was present (Tagging period, Fig. 3.2).

4.2.2 Experimental procedures

The experiments were designed and conducted by the 3S (Sea mammals, Sonar, Safety) research consortium (Miller et al 2011, Miller et al 2012, Cure et al 2013). The full CEE protocol consisted of three successive experimental phases. The first phase included randomised sonar and silent approach experiments, the second phase was dedicated to playback of killer whale sounds, and the third one was a final sonar signal transmission. A 3-4 hour baseline period was recorded after tag deployment and before the first sonar experiments. All experiments had an hour of no-exposure between them, with two hours between the first and second phase. Four whales were subject to CEE: one whale was subject only to the first experimental phase in 2008, and three whales were subject to the full protocol in 2009. In addition, in 2010, two tagged whales were subject to a playback experiment after a 8-9 hour of baseline period. This playback experiment included two sound stimuli broadcasted in the following order: a broad band noise playback as negative control and a killer whale playback. Killer whale playbacks were of unfamiliar mammal-eating killer whale sounds, simulating a potential high risk of predation (Curé et al. 2013). Broad band noise stimuli were used to ensure that whales did not react to unspecific sound stimuli and consisted of sequences from the non-calling periods of the killer whales recordings, amplified to the same RMS power as the killer whale sound stimuli.

Two main research vessels were used: an observation vessel (29m MS Strønstad) that tracked the tagged whale and a source vessel (55m R/V H.U Sverdrup II) that was used to tow the sonar source. A smaller tag boat was launched from the vessel for tagging procedures (Section 3.2.1) to take identification photographs, and for conducting killer whale playbacks. During the CEE and silent approach of the vessel, the source vessel approached the subject from a range of 6-7 km (from ahead of the whale or from the side). The speed and course of the source vessel was adjusted towards the whale, but was fixed once it reached 1 km distance and continued for 5 min after passing the whale.

The sonar source 'Socrates' (TNO, Netherlands) was towed but not transmitting during silent control approaches. During signal approaches, one of three types of sonar was transmitted: 1) MFAS-UP hyperbolic upsweep at 6-7 kHz, 2) LFAS-UP hyperbolic upsweep at 1-2 kHz, and 3) LFAS-DS hyperbolic downsweep at 1-2 kHz. Source levels were increased over the first 10 min of the exposure ('ramp-up'). The source rms levels re: 1 µPa at 1 m were 152 to 214 dB for LFAS and from 158 to 199 dB for MFAS. All sonar signal types were 1 s in duration and transmitted at 10 s intervals during ramp up and 20 s intervals during full power. The order of MFAS-UP, LFAS-UP and silent approaches was randomised for the first experimental phase. LFAS-DS was played after the killer whale playbacks in the second experimental phase. Further details about the experimental protocol can be found in Miller et al (2011 and 2012).

Received levels of the sonar signals and source-to-range were estimated and made available to this thesis chapter by the courtesy of the 3S project (Miller et al 2011). Both sound pressure level (SPL dB re: 1 µPa) and cumulative sound exposure level (SEL dB re: 1 μPa^2s) were measured. Because of the time-varying nature of the tonal sonar signal, a maximum SPL of a 200 ms running average window on the instantaneous mean-squared pressures was used (SPL_{max} dB re: 1 µPa). SEL was defined as the cumulative sum-of-square pressures, and measured the cumulative exposure over each experiment (Miller et al 2011).

All playback experiments (killer whale sounds and broad band noise control) were conducted with a loudspeaker (frequency response at 0.2-20 kHz) deployed at 8 m depth from a drifting tag boat. The boat was positioned at approximately 800 m range to the subject (estimated with

a laser range finder) when the whale was at the sea surface and after at least two hours elapsed since the last experiment. Each playback stimulus lasted 15 min. Three different killer whale recordings were used to prepare the stimuli to avoid pseudoreplication (Mc Gregor et al 1992). A hydrophone was deployed to estimate playback source levels (145-151 dB re: 1µPa at 1 m). The playback was started when the whale started a dive, immediately after the whale raised its flukes. Noise control playbacks were played first and followed by the killer whale playbacks that started at the end of the next surface phase. Thus, both stimuli playbacks were conducted on successive dive cycles. See Curé et al (2013) for a comprehensive description of the playback method.

4.2.3 Processing of incidental sonar signals

Three whales that were simultaneously tagged on 18 July, 2005 (sw05_199a, sw05_199a and sw05_199c) were exposed to incidental sonar produced by an unidentified source at an unknown location. Incidental sonar pings were monitored aurally using two-channel DTAG audio (96 kHz) and marked for start and end time (at least 200 ms in duration) using the Adobe Audition spectral frequency display. If the signal was detected aurally but its frequency range was masked by surface sounds and sperm whale clicks, it was marked for start time only and discarded for received level analysis. Many of the sonar pings contained reverberation, with up to two additional pulse arrivals (Fig. 4.2 b). In these instances, the first pulse arrival was selected for analysis unless it was masked or was lower in energy than the following one or two pulses, in which case the highest energy pulse was selected.

The analysed pulse was band-pass filtered (6^{th} order Butterworth) with frequency cut offs at 4500 and 5400 Hz, and its amplitude corrected for hydrophone sensitivity (-188 dB re 1µPaV^{-1}). The sound pressure level averaged over the 90% energy window $SPL_{90\%}$ in dB re 1µPa was measured within the filtered signal as:

$$SPL_{90\%rms} = 10\log_{10}\left(\frac{1}{T}\int p^2\,(t)dt\right) \qquad Equation\ \ 4.1$$

Where time window T is defined as the portion of the marked signal window that contained 90% of its energy ($\int p^2$). The maximum $SPL_{90\%rms}$ of the two channels was used in the analyses.

4.2.4 Time-series modelling of behavioural effects

As in Chapter 3, three response variables were modelled for both incidental and experimental exposures: 1) $state$, multinomial response variable for behavioural state, 2) $buzz$, Bernoulli presence/absence of terminal echolocation buzz as a proxy of foraging rate and 3) $ODBA$, Gamma variable as a proxy of locomotion cost.

A set of baseline explanatory variables was chosen a-priori. Probability of $state$, given previous state, was modelled by including previous state ($prevState$) as factor covariate. $State$ was used as a factor covariate in models for $ODBA$ and $buzz$ in order to allow for mean differences in the energetic proxies across the behavioural budget. Binomial models for $buzz$ were fitted to the subset of data that included foraging states (descent, layer-restricted search (LRS), and ascent). Tag id ($whale$) was used as a factor covariate in generalized linear models (GLMs, function multinom in r package nnet for $state$, and glm function in r package stats for $buzz$ and $ODBA$) and as a cluster variable in generalised estimating equations (GEE, SAS 9.3 procedure 'genmod').

GLMs were used to test which combination of experimental and exposure effects best explained the response data ($state$, $buzz$ and $ODBA$) based on the Akaike Information Criterion (AIC). This initial model selection did not account for any serial correlation in the time series, and was therefore likely to retain more effects than could be supported if data were not independent. To obtain more robust standard errors, the best AIC model was refitted within GEE which estimates the empirical correlation within each random effect ($whale$). $whale$ was included as a random effect rather than a factor covariate in order to estimate population averaged parameters. The GEE model was fitted using backwards selection where at each step, one of the coefficients was tested against the null hypothesis that it was zero (Wald test based on empirical standard error estimates) and discarded from the model if the p-value exceeded

0.05. The procedure was repeated until all remaining explanatory variables were tested $p < 0.05$ (backwards model selection).

CANDIDATE EXPLANATORY VARIABLES FOR INCIDENTAL SONAR

Potential effects of incidental sonar in 2005 were assessed first. Data were excluded from all experimental exposures and post-exposures. Pre-exposure periods for incidental sonar were not available because on all three tags incidental sonar pings were detected during or immediately after the tagging period (influence by tag boat presence, Section 3.4.3). Different post-exposure window sizes and "times since last incidental sonar ping" were therefore included to capture potential recovery to a baseline for these whales, as well as pre-exposure data from other tags as additional baseline data. Three time windows were chosen to describe exposures to incidental sonar: 10, 15 and 30 min. 10 and 15 min were used as the minimum window sizes to reflect the observed duration of bouts of the sonar transmissions and duration of gaps in reception when the whales were at surface. A 30-minute window was used to capture all time periods between consecutive bouts (Fig. 4A.1). These window sizes defined six alternative covariates: presence/absence of pings in the last 10, 15 or 30 min (*SON05_10*, *SON05_15* and *SON05_30*, respectively) and the maximum $SPL_{90\%}$ in the last 10, 15 or 30 min (*SPL_max_10*, *SPL_max_15* and *SPL_max_30*, respectively). Since all three whales that were exposed to incidental sonars were tagged during the same day, but baseline whales were tagged on other days, a day effect *SON05_w* was also included, which was set to one for the three exposed whales and zero for all other whales. For all whales that were not exposed to sonar, minutes since last ping *minFromSON05* was set to 900. The 900 min (= 15 h) was selected to reflect the typical temporal window of a DTAG record, and to exceed the maximum observed time since last ping for the three incidentally exposed whales (827 min) (Table 4.2)

CANDIDATE EXPLANATORY VARIABLES FOR EXPERIMENTAL EXPOSURES

For analysis of experimental exposures, data were excluded from incidental sonar periods (*SON05_30*), but post-exposure data periods for those records were included as no effects of incidental sonar were identified (See section 4.3.3). To test the sensitivity of results to inclusion

of these post-exposure data, selected effects models were re-fitted excluding data from each whale at a time (n=12) as well as excluding data from all the three incidentally exposed whales, and re-assessed for the importance of terms.

Candidate covariates for session and signal exposure effects were designed to capture variability in behaviour both during and after experimental sessions (*BoatPass* and *OrcaPB*) and any additive effects of signal-specific exposures (*MFAS*, *LFAS* and *OrcaS*), respectively (Table 4.2). Because LFAS downsweep transmissions were only conducted twice, LFAS upsweep and LFAS downsweep were considered as one signal type *LFAS* as they were in the same 1-2 kHz frequency band.

Session covariates were included to explain any variability due to the experimental set up alone, independent of signal exposures, such as order and source vessel approach. Sonar and killer whale playback experiments were assigned separate session covariates because of their order within the full CEE and different exposure protocols (large approaching vessel vs. stationary small vessel; 30 min exposure and 1 hour post-exposure vs. 15 min exposure and post-exposure determined by dive cycle) (Section 4.2.2). Session covariates encompassed both signal and control sessions, because the experimental design followed same protocol in both conditions. For example, presence/absence of sonar vessel approach (*BoatPass*) was set "present" during MFAS, LFAS and silent control approaches, and "absent" otherwise (Table 4.2). In the absence of any sonar signal-specific effect during signal exposure (*MFAS* or *LFAS*) therefore, session covariate *BoatPass* could capture responses to the mere approach of the vessel.

Alternative hypotheses for dose were considered for both during and post experimental sessions and signal exposures. All effects were considered a presence/absence factor for during the experiment, i.e. the sonar vessel approach (*BoatPass*), killer whale signal or noise playback (*OrcaPB*), and exposures to the three specific types of signal within BoatPass (*MFAS, LFAS*) and OrcaPB (*OrcaS*). Three different sound exposure metrics for *MFAS* and *LFAS* were compared: source level *SL*, received sound pressure level *RL* (*SPLmax*) and cumulative sound exposure level *SELcum* (Table 4.2).

For sonar vessel approaches, an order effect (*BoatPass#*) allowed any responses to sonar vessel approach to increase or decrease with repetition. *OrcaPB* could not be considered with any order or post-experiment effects, because noise controls were always conducted before playback of killer whale sounds. For all other effects, two types of post exposure were defined: a response level that would stop after a fixed time period since last exposure (*BoatPass_win*, *MFAS_win*, *LFAS_win* and *OrcaS_win*) or a decaying response as a function of time since last exposure (*minFromBoatPass*, *minFromMFAS*, *minFromLFAS* and *minFromOrcaS*). Three types of decay functions were explored: linear (for *ODBA*), exponential (due to link function for probability of *state* and *buzz*), and a quadratic by taking the square of time (min^2) (Table 4.2).

MODEL SELECTION FOR EXPERIMENTAL EXPOSURES

Exposure effects were compared in AIC model selection before assessment of post-exposure effects. First, GLMs were fitted to data that included baseline periods (excluding tagging and *SON05_30* periods) and experimental sessions (including control and signal periods, but excluding all post-experiment periods and periods between the experiments). All combinations of session effects (*BoatPass* or *BoatPass#*, and *OrcaPB*) and signal exposure effects (*MFAS*, *LFAS*, *OrcaS*) were fitted to these data and compared based upon AIC. Also *SON05_w* was included in this model search to allow for any post-exposure level effects of incidental sonar. As an alternative sonar effect, received sound pressure level (SPL) was included in the models as an interaction term with the MFAS (*MFAS:maxRL*) or LFAS (*LFAS:maxRL*) signal types, or as an effect independent of signal type (*maxRL*). If any sonar effects were retained in the lowest AIC model, the model was re-fitted with the two alternative sound metrics (*maxSL* and *maxSELcum*). The simplest (least parameters) model within two AIC units from the lowest AIC model was retained for further analyses.

To assess any post-exposure effects of retained covariates, the selected model was re-fit to all baseline, exposure and post-exposure data and checked for AIC against the baseline model. Appropriate window lengths for the different exposure covariates (*BoatPass_win*, *MFAS_win*, *LFAS_win*, *OrcaS_win*) were then determined for each included effect by model selection. The model was re-fitted with progressively increasing window size for each effect

while keeping other effects fixed. AIC was recorded for every 1-minute increase in window size up to 60 min post-exposure. Due to the large number of fitted models (i.e., 60) model averaging was used to select the most appropriate window size. A confidence set of models was selected based upon evidence ratio cut-off ($\exp(-0.5\Delta AIC) > 0.05$) suggested by Burnham and Anderson (2002). The post-exposure window duration was calculated as the Akaike weighted mean (Burnham and Anderson 2002) of the minutes in the time window. The newly defined post-window covariates were then compared to models without any post-exposure effects, and with alternative time-decaying post-exposure covariates (Table 4.2). The simplest (least parameters) model within two AIC units from the lowest AIC model was retained for the final analysis using GEE.

4.2.5 Estimating overall effects on time and energy budgets

The multinomial state models tested for instantaneous Markov changes in transition probabilities, which do not necessarily translate to overall changes in time budgets. Therefore, state-budget estimates from the multinomial models were only used as a diagnostic tool (Appendix 4C). Instead, two randomisations of baseline data were carried out in order to test for changes in time budgets given the exposure durations, sampling design and state uncertainty (Appendix 4D).

The objective of the first randomisation was to test how unusual the exposure time budgets were compared to baseline variability in time budgets across the sampled individuals (n=9, excluding incidentally exposed whales, and all experimental exposure and post-exposure periods). Baseline time budgets were sampled with replacement 10000 times. Each time, a short baseline data period ('pseudo-exposure') was sampled from a randomly selected whale (n=9, excluding incidentally exposed whales). Duration of the sample could not be less than 80% of a target duration (e.g. 30 min exposure) when the sample period exceeded the extent of the baseline data. The target duration was set to reflect the duration of each exposure of interest. To account for uncertainty in the state estimates, posterior state sequences (n=4002) were sampled with replacement. The proportion of time the whale was in each state was

calculated for each random sample, and represented as a distribution (Fig. 4D.1). The procedure was repeated for increasing the target duration from 15 to 90 min at 15 min intervals. The baseline distributions (50, 80 and 95% quantiles) could then be compared with the actual exposure time budgets at the relevant exposure durations (Fig. 4D.2). Actual exposure time budgets were also re-sampled 10000 times for posterior state sequences.

The objective of the second randomisation was to test how much exposure time budgets changed from pre-exposure periods compared to baseline changes across all baseline tag records. A 90 min pre-exposure baseline period was defined at the start of the baseline record (immediately after the tagging period). Pseudo-exposures were then drawn from the subsequent baseline record. Only whales with at least 5 hours of baseline data were used to draw pseudo-exposures (n=5: sw05_196a, sw09_153a, sw10_147a, sw10_149a and sw10_150a). As for the first randomisation, 10000 sample periods were drawn with the target duration. Each time, a posterior sequence was drawn for the randomly selected whale, and a time budget was calculated for the pre-exposure baseline and pseudo-exposure period. The pre-exposure time budget was subtracted from the pseudo-exposure budget to give a difference in the proportion of time for each state for each random sample. The resulting distribution of differences was then compared to differences in time budgets between the pre-exposure baseline and actual exposure periods. As in the first randomisation, actual exposure time budgets were also re-sampled 10000 times for posterior state sequences.

In order to estimate overall changes in energetic proxies, any effects retained in model selection were used to model probability of buzzing and ODBA excluding state effects in a GEE (Appendix 4E). Only baseline whales that were not exposed to incidental sonar, pre-exposure baseline data and the relevant experimental exposures were included in the analyses for time and energetic budgets.

4.3 Results

4.3.1 Data

A total of 165.5 hours of tag data were analysed (excluding tagging periods that were analysed in Chapter 3). 73.8 hours of pre-exposure baseline data were collected for nine whales. Additionally, 43.8 hours of data from three tag records (sw05_199a, sw05_199b and sw05_199c) that were incidentally exposed to unidentified sonar for 8.7 hours were analysed (Table 4.1). Sonar experiments on four tags (sw08_152a, sw09_141a, sw09_142a, sw09_160a) included two silent control approaches by the vessel towing the source with no transmission, six LFAS (including four up-sweeps and two down-sweeps) and four MFAS transmissions. Killer whale sound playbacks were conducted on three of the whales that were exposed to experimental sonar and on two additional whales in 2010 (sw10_149a, sw10_150a) for which control playbacks of white noise in the same frequency band were also conducted.

4.3.2 Characteristics of the incidental sonar

Incidental sonar pings were detected on tag records from three simultaneously tagged whales. The sonar was received at regular 29.5 s intervals in bouts of 6-22 pings (Fig. 4.1). Each received cycle in tags sw05_199a (n=8 bouts) and sw05_199b (n=6) consisted of a maximum 10.4 min bout of sonar (mean duration 7.3 min) and a minimum of a 25.0 min interval between consecutive bouts (mean duration 30.8 min). Only two bouts of sonar pings were detected within tag record sw05_199c: 14 pings over a 10.4 min bout in the beginning of the record, and after 2.9 hours, 13 more irregularly spaced pings over a bout of 26.7 min. Sonar pings in this bout were also received at similar intervals (min 25.3 s), but with large apparent gaps of 3.2, 6.2 and 11.3 min. This was the last bout of sonar pings detected within any of the three tag records.

Incidental sonar pings were 5100 - 4750 Hz down-sweeps that appeared to be transmitted with a 0.15 s duration (Fig. 4.2). 7 out of 222 pings were completely masked by tagged whale clicks, and 30 pings had estimated received $SPL_{90\%}$ below 95 dB re: 1µPa. The median received $SPL_{90\%}$ of the remaining 83% of pings was 113 and maximum 134 dB re: 1µPa. Once tags sw05_199b and sw05_199c were deployed, all sonar bouts that were detected on tag

sw05_199a were also detected on sw05_199b (n=5) and sw05_199c (n=1), except one bout that appeared to end before sw05_199b left the surface. Assuming a constant source level for the sonar and geometrical spreading, the source-to-whale range appeared to increase for all whales, with all whales receiving the final detected bout at a lower $SPL_{90\%}$ level. 21.1 minutes later, six lower level pings were detected at sw05_199b even though they were not detected at the other two tags, despite the fact that they were not at the sea surface at the time. The final bout of sonar pings detected at sw05_199c was not detected on either of the other two tags (Fig. 4.1).

4.3.3 Effects of the incidental sonar

There was little apparent difference in time allocation between incidental sonar periods and baseline periods (Fig. 4.3). Within the foraging states, individual average probability of buzzing within 1 min time blocks was higher during *SON05_30* (24.2% SD 10.6) compared to baseline (15.6% SD 14.8).

There was no evidence for change in state transitions during incidental exposures to sonar. *SPL_max_15* was retained in as the lowest AIC model for state, but the model only improved the baseline model (*prevState* + *whale*) by 1.18 AIC units, below the ΔAIC threshold of 2 (Fig. 4.4). GEE Wald tests did not support inclusion of *SPL_max_15* in binomial models for layer-restricted search (χ^2=0.09, p=0.76) or silent active state (χ^2=-1.52, p=0.13).

SON05_10 was retained in models for *buzz* in the AIC model selection (ΔAIC -5.2, Fig. 4.4), however the effect was only weakly supported when accounting for serial correlation in a GEE (Wald test, χ^2=3.61, p=0.057). The odds of buzzing were estimated to increase by a factor of 1.74 ($e^{0.55}$) when pings were detected in the last 10 min (Table 4.3). However, sw05_199c was likely influential in the fit, with 100% observed buzz presence in a single exposed descent (n=4 min) and layer-restricted search states (n=12 min) (Fig. 4A.2).

There was no evidence for a change in *ODBA*, given a state, as a function of incidental sonar exposures. Although *minFromSON05* was retained in the lowest AIC model (ΔAIC 42.1, Fig. 4.4), it was not supported in the equivalent GEE (Wald test, χ^2= -1.11, p=0.27). *SPL_max_30*, the

lowest AIC effects model for ODBA excluding *minFromSON05*, was not supported in a univariate GEE either (Wald test, χ^2=1.32, p=0.25).

4.3.4 Effects of the experimental sonar exposures (CEE)

DESCRIPTIVE STATISTICS BY INDIVIDUAL

Pre-exposure vs. exposure time budgets are shown for each individual in Table 4.4. Compared to pre-exposure baseline, all three exposed whales increased time spent in the silent active state during *LFAS* (from 0-1.2 to 3.3-60.0 %) and two out of five whales during *OrcaS* (from 1.7 to 81.3 %, and from 1.2 to 60.0 %). Time spent in silent active state did not increase from baseline during any other experiment (*MFAS*, silent pass or killer whale controls, Table 4.4).

The expected probability of buzzing and mean ODBA was predicted based upon pre-exposure data means within each state (Table 4E.2) and observed state budgets during exposures. All whales except sw09_141a produced fewer buzzes per minute than expected by observed state budget, or did not buzz at all, during *LFAS* and *OrcaS* exposures. State budgets explained 69.8% (linear fit R^2) of the observed probability of buzzing during *LFAS* and *OrcaS* exposures. Lower than expected probability of buzzing was also observed during all sonar and killer whale controls, but not during *MFAS* (Table 4.4).

CEE EFFECTS ON STATE TRANSITIONS

The individual average percentage of time spent in the silent active state was 7x higher during *OrcaS* (36.4% SD 36.2) and nearly 10x higher during *LFAS* (51.6% SD 30.0) compared to baseline periods (5.2% SD 13.0). Conversely, whales spent on average less than half the time in layer-restricted search (LRS) state during *OrcaS* (19.0% SD 24.8) and *LFAS* (12.2% SD 20.7) compared to baseline (47.8% SD 18.8). Whales also spent less time at the surface during LFAS (10.5% SD 6.9) than baseline (19.0% SD 4.9). Time budgets during *MFAS* (1.2% SD 1.8 in silent active, 52.5 % SD 16.5 in LRS, and 13.5 % SD 6.7 in surface state) and *Silent Pass* (3.8 % SD 5.4 in silent active, 61.8 % SD 6.9 in LRS, and 15.4 % SD 2.1 in surface state) did not appear to differ markedly from baseline (Fig. 4.5).

AIC selection for *state* retained *LFAS* and *OrcaS* in models for baseline + exposure data, improving the baseline model (*prevState* + *whale*) by 62.2 AIC units. Including post-exposure data, *LFAS* and *OrcaS* improved the baseline model by 60.4 AIC units. There was support for increasing the window length of the effects: replacing *LFAS* with *LFAS_8* and *OrcaS* with *OrcaS_19* (Fig. 4B.2) improved the model by further 9.2 and 27.7 AIC units, respectively (Table 4.5, Table 4B.1). Also $min^2FromOrcaS$ was selected, bringing the total ΔAIC compared to baseline model to 126.1 units.

Silent active state and the foraging states (descent, LRS or ascent) were estimated to have had the greatest change during LFAS and orca playbacks, and were therefore used as binomial response variables for GEE (presence/absence of silent active state, and presence/absence of one of the three foraging states, respectively). In the resulting GEE models, $min^2FromOrcaS$ was not supported as an explanatory variable neither for silent active state transitions (Wald test, χ^2=2.9, p=0.08) nor transitions to one of the foraging states (χ^2=0.3, p=0.56) and was excluded from further GEE and GLM models. In both GEE models for silent active state and foraging state, *LFAS_8* and *OrcaS_19* were supported strongly with or without $min^2FromOrcaS$ (GEE Wald test p < 0.006, Table 4.6).

The multinomial model for state (*LFAS_8* + *OrcaS_19*) provided a reasonable fit to data, fitting 86.8% of all states correctly, 82.3% correct during LFAS_8 and 80.4% correct during OrcaS_19. The model fit best to state 3 (93.2% of correct states) and worst to state 6 (69.2% correct states) (Fig. 4B.3). Only weak positive and negative serial correlation was detected in the residuals of multinomial model that did not explicitly model autocorrelation (Fig. 4B.5). The binomial GEE models for silent active state and foraging state appeared to fit data well (Figs. 4B.5-6). Both models retained significance of identical terms when data were excluded from the three incidentally exposed whales (post *SON05_30* periods) (Wald test, p-value < 0.008). However, cluster-wise Cook's distance identified sw09_141a as influential in the model for silent active state (Fig. 4B.6). The multinomial coefficient estimates for *LFAS_8* and *OrcaS_19* were not changed significantly when fitted excluding data from each individual or all three incidentally exposed whales, and excluding sw09_141a only appeared to increase the magnitude of the effects on state 6. All full

models also retained ΔAIC < -67.6 to the baseline model (Fig. 4B.4). The influence of sw09_141 was therefore likely related to the baseline probability of mean transition to silent active state across the whales, rather than the uncertainty in the effects per se.

In the binomial GEE for foraging state, the odds of transition to descent, LRS or ascent was estimated to decrease by a factor of 0.41 (95% CI [0.27, 0.61]) during *LFAS* exposure up until 8 minutes after exposure, and decrease by a factor 0.40 (95% CI [0.21, 0.76]) during *OrcaS* exposure up until 19 minutes after exposure. Conversely, the odds of silent active state were estimated to increase by a factor of 6.51 (95% CI [3.3, 12.7]) during *LFAS_8*, and increase by a factor 5.73 (95% CI [3.6, 9.2]) during *OrcaS_19* based on the binomial GEE (Table 4.6). The respective multinomial model coefficients for silent active state were similar for *LFAS_8* ($e^{1.88+/-2xSE}$= [3.5, 12.0]) and slightly lower for *OrcaS_19* ($e^{1.65+/-2xSE}$ = [2.4, 11.3]) (Table 4B.2).

CEE EFFECTS ON PROBABILITY OF BUZZING, GIVEN A STATE

Within the three foraging states (descent, LRS and ascent), individual average buzz rate (% presence in 1-min bins) was less than half during *LFAS* (7.6%, SD 9.6, n=128 min), *OrcaS* (7.1%, SD 9.5, n=44) and *Silent pass* (3.7%. SD 0.6, n=80) than during pre-experiment baseline periods (19.8%, SD 10.1). The difference was smaller for *MFAS* (16.4% SD 6.4, n=193 min) and *OrcaPB* (11.4%, SD 6.2). Compared to pre-exposure baseline within each whale, an overall decrease in buzz rate was observed for all four whales during *LFAS* (36.9-100% decrease; two whales ceased buzzing completely during exposure), and four out of five whales exposed to *OrcaS* (45-100% decrease; three whales ceased buzzing completely during exposure).

AIC selection for *buzz* retained *BoatPass*, *OrcaPB* and *MFAS*, which improved the baseline model by 28.6 AIC units when including post-exposure data. There was no support for increasing window size beyond the duration of the exposure for *BoatPass* or *MFAS*, (Fig. 4B.2 b), while the inclusion of post-exposure effects *minFromMFAS* and $min^2FromBoatPass$ improved the model by a further 5.3 AIC units (Table 4.5). Post-exposure effects for *OrcaPB* were not tested for due to confounding effects of playback order.

The Wald tests for GEE model for buzz supported *BoatPass*, *OrcaPB* and *MFAS* ($p<0.001$), but not *minFromMFAS* ($\chi^2=0.6$, $p=0.45$) or $min^2FromBoatPass$ ($\chi^2=0.7$, $p=0.41$). The GEE model was

re-fitted without *minFromMFAS* first and re-evaluated for Wald tests. *minFromBoatPass* gained no support (χ^2=0.5, p=0.49) and was dropped from the final model for buzz. In the final GEE model for *buzz*, there was strong support for *BoatPass*, *OrcaPB* and *MFAS*, including or excluding post *SON03_30* data (Wald test p-values ≤ 0.005, Table 4.5).

The final GEE model appeared to be a good fit to the data, with little influence of data or clusters on model fit. The GEE model residuals were positively correlated up to lags of 60 min (Fig. 4B.8).

Odds of buzzing was estimated to decrease during *BoatPass* by a factor of 0.28 (95% CI [0.17, 0.46]) in a given 1-min bin. Odds of buzzing was also estimated to decrease during *OrcaPB*, but to a lesser degree (by a factor of 0.60 [0.42, 0.86]). Relative to LFAS exposures and silent approaches, odds of buzzing was estimated a factor 2.69 higher during *MFAS* (95% CI [1.52, 4.76]) (Table 4.7).

CEE EFFECTS ON ODBA, GIVEN A STATE

There were no obvious overall trends in ODBA across states in response to CEE exposures (Figs. 4.5-6). Individual average ODBA in ascent and surface states appeared to have been slightly elevated during exposures to LFAS. Within the four LFAS exposed whales, increase in mean ODBA was observed only in 2 whales for ascent states (20.1 % and 97.3% increase), and 3 whales for surface states compare to pre-exposure (by 23.7-63.7%). Within the five OrcaS exposed whales, increase in mean ODBA was observed in 2 whales for ascent states (24.0 % and 71.0% increase), and 3 whales for surface states (43.1-51.6% increase).

AIC selection for *ODBA* retained *BoatPass#*, *MFAS_13* and *LFAS* (Table 4B.1, Fig. 4B.2 c), however there was weak support only for *LFAS* in the final GEE analysis. Wald test did support *BoatPass#* when *MFAS_13* and *LFAS* were also included in the GEE (χ^2=9.3, p=0.002) (Table 4.5). However, *MFAS_13* was not supported by Wald test (χ^2=2.1, p=0.15) and was therefore removed from the full model. In turn, *BoatPass#* was not supported in the resulting model (χ^2=1.1, p=0.30), and excluding it reduced support for *LFAS* above the 5% significance level (χ^2=2.8, p=0.094). Removing *LFAS* from the model, *minFromLFAS* retained weak support (χ^2=3.9, p=0.049), but the effect did not hold when excluding data from the incidentally

exposed whales (χ^2=0.9, p=0.33). Both coefficients for *LFAS* and *minFromLFAS* were estimated positive; the 95% confidence intervals of the coefficient for *LFAS* encompassed zero [-0.51, 6.47], while the coefficient for *minFromLFAS* was estimated very small (95% CI [0.30, 0.65].

4.3.5 Effects on overall time and energetic budgets

Only *LFAS* and killer whale sound playbacks were considered in the analysis for overall time and energetic budgets because there was little evidence for a change in state transitions or proxies, given a state, in response to other signals. Data were included for baseline whales that were not exposed to incidental sonar, pre-exposure baseline and the relevant exposures (*LFAS_8* and *OrcaS_19*).

Uncertainty in baseline state budgets appeared to stabilise when resampled time periods exceeded ~45 min in duration (Fig. 4D.2 a-b). Time spent in LRS state was the most variable across the baseline pseudo-exposures, with 95% of the proportions falling between [.00, .80]; in contrast, the upper limit of the quantile was less than 0.40 for surface, descent and ascent states (at 45 min target window) (Figure 4D.2 a). Nevertheless, all individuals except sw09_160a spent less than the baseline average time in LRS during both exposures, with most exposures in the lower 10% quantile at the relevant time windows (35-69 min). Similarly due to high baseline variability, time spent in the silent active state was near but did not exceed the 95% quantile for most experiments. Silent active state was within the 80% quantile of proportions only for four experiments (from two whales sw09_160a and sw10_150a). Surface time was also reduced during 5 of the experiments (3 whales; 2 *OrcaS* and 3 *LFAS*) where the proportion of time was 2.5% or less of the baseline pseudo-exposure probability mass. In all *LFAS* and *OrcaS* experiments, the proportion of time spent in descent, ascent and drifting states were well within the typical (95%) pseudo-exposure time budgets sampled from the baseline behaviour (Figure 4D.2 a).

There was a significant increase in the proportion of time (+0.2 - +0.7) spent in silent active state during most of the experiments (at two-tailed 5% level, 3 out of 5 *OrcaS*, and 5 out of 6

LFAS exposures) compared to pre-exposure baseline (90 min following tagging period) (Figure 4D.2b). Increase in silent active state was within the 95% quantile of baseline samples for the first experiment of sw09_160a, but not the subsequent *LFAS* and *OrcaS* exposures to the same whale. The mean exposure budget of sw09_141a during *OrcaS* was above the 95% quantile of pseudo-exposure, but large state uncertainty precluded detection of a significant effect. Only two whales appeared to reduce time in LRS state, sw08_152a and sw10_149a. No unusual changes were found for time spent in other states (surface, descent, ascent, or resting) (Figure 4D.2 b).

To estimate the effects on energy budget, given the exposure durations, probability of buzzing and ODBA were modelled in a GEE excluding state effects. There was good support for an overall reduction in probability of buzzing during LFAS_8 (Wald test, χ^2=5.05, p=0.025) but not during OrcaS_19 (χ^2=0.39, p=0.53). Overall odds of buzzing was estimated to decrease by a factor of 0.25 (95% CI [0.21, 0.84]) during *LFAS* + 8 min into the post-exposure (Table 4E.1). There was no support for a change in ODBA during *LFAS* (Wald test, χ^2=0.92, p=0.34) and weak evidence for a reduction in ODBA during orca signal playback (χ^2=3.33, p=0.068). However, the 95% CI estimate for a reduction in ODBA during orca signal playback included zero [-4.38, 0.16] (Table 4E.1).

4.4 Discussion

This chapter set out to estimate the potential time and energetic cost of sperm whale behavioural response to naval sonar compared to costs incurred in an anti-predator context (playbacks of killer whale sounds). Time-series modelling estimated near identical decreases in the probability of state transition to foraging between vessel approaches transmitting *LFAS* and stationary playback of killer whale sounds (*OrcaS*) (odds decreased by a factor of 0.41 and 0.40, respectively). The effect was estimated to last 8 min into *LFAS* post-exposure and 19 min into *OrcaS* post-exposure. Combined with the exposure duration (~30 and 15 min, respectively), this is about the duration of an average sperm whale dive or a short dive cycle (40 min, Watwood et al 2006). Re-sampling of baseline data with dummy exposures confirmed that

the changes in state-transitions translated to an atypical increase in time allocation to a non-resting, non-foraging state within this time period. If the behavioural state changes were persistent over repeated or longer period of LFAS exposure, sperm whale energetic budget could be affected by a reduced energy intake. Assuming that time and energy is traded off as a function of perceived risk, sperm whales perceived an LFA sonar (at 1-2 kHz and source sound pressure levels of ~ 200-214 dB re 1µPa m approaching from 10 km to a minimum approach distance of 320 m) as high a risk as detection of an unfamiliar mammal-eating predator at ranges <1km.

4.4.1 Behavioural changes and strength of evidence

Three statistical hypotheses were tested to assess different possible behavioural responses of sperm whales to naval sonar: 1) changes in functional state transitions, 2) change in probability of terminal echolocation sequence (*buzz*, proxy for foraging success), given foraging effort, and 3) change in overall dynamic body acceleration (*ODBA*, proxy for locomotion cost), given the functional time budget.

Changes in functional state transitions were found during *LFAS* transmissions and playbacks of killer whale sounds. The individual average time spent in silent active state was nearly 10x higher during *LFAS* exposure (51.6%) and 7x higher during *OrcaS* (36.4%) compared to baseline (5.2%) (Fig. 4.5). GEE analysis lent strong support for a reduced probability of transition to foraging states and increased probability of transition to silent active state during both exposures (Table 4.5). In contrast, there was little evidence that sperm whales would have modified their functional state budget during experimental exposures to *MFAS*, experimental control periods (silent approaches, or stationary playback of broadband noise) (Table 4.5) or incidental sonar exposures (Fig. 4.4).

These results suggest a functional shift in behaviour from foraging to non-foraging, which is also supported by expert scoring of cessation of feeding, change in vocal behaviour, and change in dive behaviour during the same *LFAS* and killer whale playback experiments, and lack thereof during silent transmission and *MFAS* (Miller et al 2012, Curé et al 2013). Similarly, Sivle

et al (2012) reported more consistent dive responses (shallower deep dives) to *LFAS* than *MFAS* for these whales. Avoidance responses reported by Miller et al. (2012) may be only partially captured by the functional states that describe diving behaviour and did not account for horizontal movement.

Four out of six exposed whales reduced time spent in layer-restricted search (LRS) by 66-100% during *LFAS* and/or *OrcaS* compared to pre-exposure baseline (Table 4.4). The response profiles of the two other whales (sw09_141 and sw09_160) were not as clear. sw09_141 appeared to engage in social interactions during pre-exposure baseline, with 46% of time spent in a silent active state that coincided with detection of social sounds (codas and slow-clicks) and a visual group size estimate of 2-4 whales at the beginning of the tag record. sw09_141 subsequently remained in the silent active state throughout *LFAS* and *OrcaS*, although with a foraging period in between that overlapped with *MFAS* and silent pass. In contrast, sw09_160 spent 87% of the baseline period in foraging states (descent, LRS and ascent) and had the highest baseline probability of buzzing in LRS state (58%) than any other whale in this study (Table 4E.2). sw09_160 continued foraging during most of the exposures to both *LFAS* and *OrcaS*, with only a brief transition to silent active state during the first exposure to *LFAS* (up-sweep *LFAS*; both *LFAS* down-sweep exposures were conducted after up-sweep *LFAS*) (Table 4.4, full time series in 3D.3). Although anecdotal as evidence, these two whales demonstrate two contrasting behavioural contexts - socialising vs. intense foraging at a high quality patch (indicated by a high buzz rate) - that could be important in modulating their responsiveness to anthropogenic stimuli.

The multinomial models for state appeared effective at detecting and quantifying changes in state transitions. Although the models assumed that state-transitions only depended upon the previous step (1st order Markov), good concordance was found between predicted and observed time budgets across individuals (Fig. 4C.1). Inclusion of time in the most recent state as a covariate changed fitted states by less than 2% during *LFAS* and *OrcaS* exposures (Appendix 4C). Furthermore, an independent analysis based on randomisation of baseline time budgets agreed with multinomial state estimates: both methods found the greatest change in

increased silent active state, and reduced but more variable probability of surfacing and LRS state during *LFAS* and *OrcaS* exposures (Table 4B.2, Fig. 4Da-b).

There was no clear support for a change in foraging success, given a foraging state, during any experimental signal exposure. Model selection retained *MFAS* and a session covariate for sonar transmission boat (*BoatPass*), which together captured variability in buzzing better than *LFAS* alone. Probability of a buzz in foraging state was estimated to decrease during LFAS and silent controls, and increase during *MFAS* compared to pre-exposure baseline. However, *MFAS* was not supported as a univariate variable in the model (Table 4B.1), and it was estimated as a positive effect in the full model (Table 4.7). These results do not suggest that *MFAS* have a direct effect on probability of buzzing, given a state, but rather that the reduction in probability buzzing due to *LFAS* was similar in magnitude to that during the two silent controls. While one of the silent controls was conducted before the full CEE (sw09_142a), the other control was conducted after *LFAS* and *MFAS* exposures (sw09_141a). With a brief transition to silent active state during the ascent phase in the latter control, it is possible that the whale (sw09_141a) was somewhat sensitized to the source vessel. Such sensitization would be expected if the whale learnt to associate the source vessel approach with an increased risk of sound exposure.

Similarly for killer whale playbacks, the probability of buzzing during foraging states was predicted to decrease during exposure, but the model could not differentiate between killer whale sounds and noise playbacks (Table 4.7). Buzzing only occurred during the foraging states (descent, ascent and LRS), and therefore lack of data (foraging time) during *LFAS* and *OrcaS* exposures appeared to preclude detection of significant effects on buzzing, given a state.

There was little evidence for change in locomotion cost, given a state, during experimental exposures. There may have been a small increase in *ODBA* during *LFAS*, but the effect was sensitive to the inclusion of other explanatory variables and data from the post-exposure period of incidental sonar.

Three whales that were tagged during the same day (sw05_199a, sw05_199b and sw05_199c) were exposed to unidentified sonar in the overlapping tag records (Fig. 4.1). No

pre-exposure baseline data were available because incidental sonars were detected immediately after tag-on time. Baseline data from other whales, alternative time windows and time-decaying covariates were used instead to contrast behaviour during exposures (Table 4.2). During the incidental sonar duty cycle (~10 min bout + 30 min post-exposure), the whales performed both very deep (>1000 m) and shallow (~200m) dives (Fig. 4.1), and overall spent a large proportion of time foraging and buzzing (Fig. 4.3). There was no evidence for a change in state transitions or ODBA during incidental exposures. There was weak support (p=0.057) for an increase in probability of buzzing, given a state, when pings were detected within the last 10 min (*SON05_10*). However, the effect was not supported within the total exposure period (*SON05_30* ΔAIC < 2, Fig. 4.4) and sw09_199c was highly influential in the model fit (Fig. 4A.2). The effect of *SON05_10* therefore did not appear robust in the models. It is possible that its importance may have been artificially inflated by confounding effects of dive depth on incidental sonar detection, as well as potential pseudo-replication of similar foraging conditions.

4.4.2 Implications for an energetic budget

Changes in state transition probabilities or energetic proxy indicators such as *buzz* and *ODBA* could indicate changes in the total energetic budget during an exposure. Of course, since it was not possible for us to measure actual energy quantities, it is possible that energy acquisition or usage could be compensated for after response to changes in time budget. There was little evidence of such effects in the measured proxies; although compensation could also occur by variable capture rates and calorific value of prey that could not be measured directly (Section 5.5.2). For the purposes of the present chapter, therefore, overall effects on the energetic proxies were considered as mostly a consequence of changes in functional behaviour during *LFAS_8* (exposure + 8 minutes into post-exposure) and *OrcaS_19* (exposure + 19 minutes into post-exposure).

Randomisation of time budgets from baseline data suggested that changes in state transitions indeed translated to changes in time budgets both during *OrcaS* and *LFAS*. Sperm

whales spent 50-80% of their time in silent active state and less than 20% of time in LRS state during most of the *OrcaS* and *LFAS* experiments, which was unusual but mostly within the observed range of behaviours in baseline data (Fig. 4D.2 a). However, the increase in time spent in silent active state compared to pre-exposure budgets was highly unusual (<<5%) for all but three of the experiments (Fig. 4D.2 b).

GEE excluding state effects suggested an overall reduction in probability of buzzing during *LFAS_8*, although with a wide confidence interval [0.2, 0.8]. There was no evidence for an overall change in probability of buzzing during *OrcaS_19* (Table 4E.1). These results indicate that although an overall individual-average reduced probability of buzzing could be detected for *LFAS*, the short-term reductions in probability of buzzing during orca playbacks could have been achieved by chance alone if lower foraging effort was not considered. This was expected as sperm whales spend time in functionally important non-foraging behaviours (surfacing, resting, socialising) and therefore short cessations of feeding are not unusual. The importance of accounting for all functional behaviours was further highlighted by high variability in proportion of time spent in layer-restricted search state compared to proportion of time spent in the silent active state (Fig. 4D.2). The observed changes in state budget alone explained 70% of reduction in buzz rate during both *LFAS* and *OrcaS* exposure (Table 4.4).

There was little evidence for a state-dependent or overall change in *ODBA* during exposures, suggesting that reduced intake of prey rather than increased expenditure contributed to an energetic trade-off. Whales did not switch to resting or increase drifting rates during exposures (Fig. 4.5), indicating that reduced energy intake was not immediately compensated for (Miller et al 2009). Retaining or even lowering locomotion costs would be a more adaptive behavioural response for a deep diving marine mammal that might incur physiological costs and even injury from sudden bursts of activity while performing deep and long anaerobic dives (Fahlman et al 2014). While effects on *ODBA* were not found, the unusually small proportion of time spent at the sea surface (5 experiments < 2.5% quantile, Fig. 4D.2 a) may have indicated a somewhat compromised recovery at surface and subsequent dive metabolism. Nevertheless, it should be emphasised that lack of any energetic “flight” may not indicate lack of a stress response (Wright

et al 2011), especially in a deep-diving marine mammal whose behavioural options are likely to be constrained physiologically.

In conclusion, the overall reduction in buzz rates and increased time spent in active non-foraging modes of behaviour suggest that a biologically significant reduction in prey intake could be a consequence if the behavioural changes persisted over longer or repeated exposures to either LFAS or the unfamiliar mammal-eating predator. This prediction is similar to the results of Williams et al (2006) who by assigning fixed energy expenditures to visually classified activity states found that reduced energy intake rather than increased energy expenditure was the consequence of altered time allocation in killer whale groups exposed to vessel traffic. Nevertheless, the temporal extent to which sperm whales are exposed to types of naval sonar at regional or global scales is not known, nor whether habituation/sensitisation processes would allow for subsequent chronic reductions in foraging success.

4.4.3 Comparison of response stimuli

Experimental sonars were transmitted from an approaching source, and *LFAS* was both transmitted and received at higher sound pressure levels at full power (max received SPL_{max} 169.6 dB re 1µPa) than *MFAS* (max received SPL_{max} 157.8 dB re 1µPa). Incidental sonar pings were received at the lowest sound pressure levels (max $SPL_{90\%}$ 133.5 dB re 1µPa). Incidental sonars (4.75-5.10 kHz downsweep) were closer to the fundamental frequency of *MFAS* upsweeps (6-7 kHz) than *LFAS* upsweeps or downsweeps (1-2 kHz). However, the *LFAS* sonar source produced harmonic distortions when transmitting at higher source levels (Miller et al 2011) which were prominent in the received spectral signature of *LFAS* at full power (Fig. 4.2). Individual harmonics in *LFAS* were a minimum of 15 dB less intense than the fundamental, and thus there was little difference between received SPL_{max} evaluated broadband or at the operational band. No harmonics were present in the incidentally received *MFAS* signals, although this may have been due to absorption and the greater distance from the whale to the source.

Note slight differences in the two received level measurements SPL_{max} and $SPL_{90\%}$. SPL_{max} for experimental sonars was taken as the maximum of a 200 ms running average window over the 1 s signal. Due to masking and detection constraints from an unknown source, the duration of the manually detected incidental sonar signals was variable (mean 237 ms, range [165, 553] ms). To achieve a comparable maximum SPL, the rms pressure was taken over a 90% energy window over the detected signal length. The resulting durations over which $SPL_{90\%}$ were measured ranged between 128 and 301 ms (mean 197 ms). Nevertheless, the received SPL_{max} were surprisingly consistent temporally across the three incidentally exposed tags (Fig. 4.1), indicating that the movement and sensitivity of the receivers had relatively minor influence on the received levels. It therefore seems reasonable to assume that the incidental sonars were relatively distant (tens of kilometres).

Incidental sonar was received in bouts of sonar pings that were shorter in duration (max 10.4 min) than the experimental sonar exposures (~45 min), and at slightly longer inter-ping intervals (29.5 s) than the experimental sonars (20 s). It is not known for how long these sperm whales received the incidental sonar signals before tags were attached. Nevertheless, the incidental sonar cycle was repeated eight to nine times bringing the minimum total exposure time to one hour for sw05_199a. From the synchronous timing and signal level between tags, it is possible to infer that sw05_199b was within a detectible range of the incidental sonar for a similar duration (> 4 h) (Fig. 4.1).

All whales that were exposed to incidental sonar and experimental *MFAS* spent most of their time in foraging states during those exposure periods (Table 4.4, Fig. 4.6), while transitions to foraging states were significantly reduced during *LFAS* exposures (Table 4.6) and allocation to non-foraging states was significantly increased for most *LFAS* exposures both within and across individuals (Fig. 4D.2 a-b). Switching to silent active state was observed within minutes of onset of exposure (Fig. 4.7) and both in final ascent and first descent of foraging dives (sw08_152a, sw09_142a), as well as in surface state (sw09_160a). From the same data, Miller et al (2012) also scored higher severity responses to *LFAS* than *MFAS* for sperm whales.

These results indicate that key behaviours were traded off against perceived risk during experimental *LFAS* exposures (Section 1.1). Here, the conceptual model developed in Chapter 1 (Fig. 1.1) is used to discuss potential mechanisms that might have altered the perceived cost-benefit space of foraging during experimental LFAS exposures: 1) direct physical effects - temporary threshold shift, sensitivity to sound levels 2) availability of information - information masking, 3) resources - indirect effects on prey, and 4) perceptual effects - association of different aspects of the experiment (approach angle and relative speed, signal level and spectral content) with risk.

DIRECT PHYSICAL EFFECTS

The sensation level has been defined as the dB amount that a received level exceeds auditory threshold of detection at a specific frequency band (Southall et al 2007). Unfortunately, audiograms have not yet been measured for the sperm whale in these frequency bands and therefore sensation levels could not be quantified for this analysis. Nevertheless, auditory brainstem responses (ABR) on a stranded sperm whale neonate suggest 5-20 kHz to be the most sensitive frequency region (Ridgway and Carder 2001), which coincides with spectral endpoints (5-24 kHz at -10 dB) measured from on-axis sperm whale echolocation clicks (Madsen et al 2002). If the sperm whale hearing filter indeed tapered off below 2-5 kHz, a lower detection threshold would be expected for the *MFAS* rather than the *LFAS* fundamental. On the other hand, a significantly lower sensitivity at 1-2 kHz would appear somewhat disadvantageous if this bandwidth was also important for long-range detection of conspecifics or mammal-eating killer whales. The sensitivity of sperm whales to *MFAS* would have therefore been expected to be similar if not higher than to *LFAS* at a given received level. The maximum received SPL of both experimental signals (*MFAS*: 158 and *LFAS*: 170 dB) were below 192-201 dB for 1-s and 250 ms tones that induced masked temporary threshold shifts (MTTS) in bottle nose dolphins and beluga whales (Schdlundt et al 2000) (re 1 µPa m).

Within the tested range of received level of sonar, the state transition analysis did not lend support for a particular response threshold. Presence/absence of LFAS was retained in model selection over interactions or main effects of received level (Table 4.5, Table 4B.1). LFAS

improved the baseline model by 43.5 AIC units while the sound metrics (*MFAS/LFAS* combined) improved the baseline model by 15.2-18.6 AIC units. All four exposed whales appeared to start switching to silent active state already at a cumulative SEL of 135-145 dB, which was ~20 dB below the maximum cumulative SEL of *MFAS* (median 139.6 dB, max 160.6 dB) (Fig. 4.7). These results are consistent with Miller et al (2012) who scored responses at a wide range of received levels for the same set of experiments. State transitions also appeared to be unaffected by silent approaches (Table 4B.1) or range during *MFAS* experiments (Fig. 4.7), suggesting that lower level and lower frequency source vessel engine noise alone was not directly affecting foraging behaviour.

AVAILABILITY OF INFORMATION

Sperm whale vocalisations consist overwhelmingly of echolocation clicks, and any masking of returning echoes from prey or other environment (sea floor) could have disrupted foraging states by limiting the individuals' ability to locate prey and navigate. In their foraging ground, male sperm whales may also listen to other individuals and use vocalisations such as codas in social display and interactions (Oliveira et al 2013, Curé et al 2013). However, most whales documented here appeared to be primarily solitary and foraging, with the exception of sw09_141a. The discontinuous and narrow-band sonar signals seem unlikely to have limited the communication range of the broad-band (centroid frequency of regular clicks 8-26 kHz, Madsen et al 2002) and intense (> 220 dB re 1µPa @ 1m source level) echolocation clicks (Madsen et al 2002, Møhl et al 2003) in a way that would have elicited response in a foraging context.

In contrast, echolocation on prey involves the processing of a potentially more complicated and noisy auditory scene that may vary in composure of target strengths, prey sizes and mobility (Madsen et al 2005). We can speculate that the received sonar, including associated reverberation and harmonics, may have masked portions of the received echo signatures that contain information about prey.

sw09_160a switched to silent active state at the onset of first *LFAS* (upsweep), but appeared to continue in foraging states and buzzing (n=8) throughout the second playback of *LFAS*

(downsweep) (Table 4.4). Given the nearly identical received ramp-up levels and similar potential for masking between the upsweep and down-sweep exposures (Fig. 4.2 a), recovery of foraging in a particular foraging patch appears somewhat inconsistent with the information masking hypothesis. Nevertheless, this observation is inconclusive as it relies upon a sample size of one individual, and lacks information about prey field.

RESOURCES

A primary habitat quality that could have been affected by sonar is sperm whale prey, fish and cephalopods (Miller et al 2009). While fish has been reported as major part in the diet of sperm whales in Icelandic waters (Martin and Clarke 1986), more recent studies of stranded specimens in the NE Atlantic have documented predominance of cephalopods such as Gonatus sp. (Santos et al 1999, Simon et al 2003). Most fish species can detect sounds at 50-1500 Hz, and species with swim bladders may be susceptible to injury from both high-level impulsive sounds and continuous low frequency (150 Hz) tones (Popper and Hastings 2009). Cephalopods have sensory organs called 'statocysts' that are thought to function as balance and acceleration detectors, but whether statocysts or lateral line hair cells sense acoustic particle motion and/or sound pressure is debated (Mooney et al 2012). 50-400 Hz sinusoidal sweeps at RL 157 dB 1 re 1µPa have been shown to cause injury in statocysts for a range of cephalopod species (Sepia, Octopus, Loligo, Illex sp. Solé et al 2013). On the other hand, broadband and intense (>220 dB re 1µPa pp) odontocete clicks do not appear to debilitate or trigger anti-predatory behavioural responses in *Loligo pealeii* (Wilson et al 2007). Cephalopod hearing remains poorly understood, but in the light of the current evidence, it seems unlikely that Cephalopods are particularly sensitive at the 1-2 kHz band. From an adaptive point of view, it also seems unlikely that either prey type could become unavailable to the predator at the time scales that switching to silent active state was observed, and resumed again estimated 8 min post-exposure (Fig. 4B.2 a).

PERCEPTUAL EFFECTS

If no costs or risks were actually realised, then sperm whales were most likely responding on mere association or anticipation of cost or risk (e.g. risk-disturbance hypothesis, Frid and Dill

2002). Such cue-response systems are influenced both by phenotype (past selection, ontogeny) and learning (individual or social) (Sih 2013). Military sonars have been in use since the second world war (D'Amico et al 2009), and it is possible that the incidentally detected sonars are relatively commonplace in Northern Norway and much of the world's oceans. In 2003, ~ 300 MFA sonars were in use globally (Watts 2003). Learning may also occur in the course of the exposure (habituation/sensitisation). Incidental sonars were detected immediately after tag-on time, suggesting that at least two of the whales were exposed to the sonar already before tag deployment (Fig. 4.2). No data was available on how long the three whales may have been exposed to the incidental sonar previously. Although comparison of experimental and *LFAS* and *MFAS* are suggestive of the importance of signal type (spectral signature and/or source level), the data does not allow conclusion on whether an initial response to the incidental sonar could have been similarly strong to the experimental *LFAS* exposures.

More direct and fast approaches of threats have been shown to increase perceived risk and subsequent responsiveness in a range of species (Frid and Dill 2002, Stankowich and Blumstein 2005, Domenici et al 2011). Similarly here, experimental *LFAS* and *MFAS* were transmitted on a controlled distance approach to focal whales, and may have triggered a stronger response as a result compared to incidental sonar, which did not perform a close approach. One explanation for the lower estimated recovery time in response to LFAS (8 min) compared to OrcaS playbacks (19 min) (Fig. 4B.2 a) could be the fact of having a moving vs. a stationary source respectively. Whales would have been able to monitor source vessel departure. Greater distance to the source may have also modulated lower responsiveness to the incidental sonar, compared to CEE exposures. DeRuiter et al (2013 a) found that while two beaked whales (Ziphius cavirostis) showed intense and energetically expensive responses to experimental MFA below regulated levels at 89-127 dB (rms re 1µPa), behavioural changes were not detected in an individual that was exposed at comparable levels (78-106 dB) to distant (>100 km) incidental MFA sonar. The lowest frequency naval sonars can be detectable at 100s of km (Hildebrand 2009).

Miller et al (2012) suggested that whales could have responded to the higher source level used in the *LFAS* experiments by integrating information about the distance to the source. Combined with monitoring of approach, the response may have been a flight response from the course of the approaching loud source, perhaps to avoid anticipated exposure. Due to relatively short ramp-up of *LFAS* and higher source levels at full power (Fig. 4.7), presence/absence of higher source level may have simply explained more data than a continuous escalation of source level.

The lower-frequency spectral signature of *LFAS* itself could have been associated with an increased risk, as it might indicate a larger source (lower frequency ~ larger resonating space) or a more specific 'ghost of signal past' (Sih 2013), such as a predator. The frequency-size hypothesis may not hold if whales integrated other information than frequency (such as reverberation, vessel approach), which seems likely for cetaceans that rely on hearing for most sensory input from their environment. The killer whale is a cosmopolitan species with localised ecotypes and it is possible that both the phenotypic and evolutionary diversification of the mammalian predator (Bruyn et al 2013) has imposed selection for sperm whales that are wary of sounds at the frequency range of their vocalisations. Killer whale vocalisations occur across a frequency range of 0.5-120 kHz, but significant for conspecific eavesdropping, energy can be found at the 1-2 kHz band (Cure et al 2012). Outright misidentification of *LFAS* signal as a vocalisation of a mammal-eating predator, e.g. due to similar harmonics (Zimmer and Tyack 2007) is not fully supported by data that shows extended post-exposure effects for orca signal playbacks (Fig. 4B.2 a).

Unlike motor vessels, which produce sound continuously while underway, killer whales can move silently, so whales have little means to monitor their movements outside their visual perimeter. Mammal-eating killer whales cease vocalising before engaging in cryptic hunting strategy (Deecke et al 2005). Thus, the ceasing of predator vocalisations itself could be a signal that increased the perceived predation risk and subsequent anti-predatory response in the post-exposure period of the playbacks. It seems likely that sperm whales assimilated information about the acoustic signature of the sonar pings and vessel noise to differentiate

between sonar transmission and killer whale playbacks to produce a slightly extended response to killer whale playbacks.

4.4.4 Conclusions

Given the relatively short duration ramp-up and small sample size, there was little data to contrast effects of spectral signature and source level as separate disturbance stimuli. Nevertheless, sensitivity to higher source level or the specific 1-2 kHz frequency band with harmonics remain plausible factors explaining why sperm whales were more disturbed by 1-2 kHz LFAS sonar used in this study than the 6-7 kHz *MFAS* sonar or the 4.7-5.1 kHz incidental sonar signals. Perceptual-contextual effects were likely to play a role. Antipredatory and defence behaviours to specific and generalised threatening stimuli are likely based on similar response mechanisms, and such behaviours may have been selected in response to signals whose recognition ranges from uncertain to highly certain (Sih 2013). For example, similar to reported effects of sonar (DeRuiter et al 2013 a), mammal-eating killer whale playbacks elicited prolonged avoidance in a Blainville's beaked whales (Allen et al 2014). Future research on sperm whale behavioural responses could address whether specific antipredatory behaviours such as initiation of social sounds (Cure et al 2013) are also used in the context of novel stimuli, and how much time and energy is traded off in such antipredatory modes of behaviour.

Buzz rates were highly variable across individuals and contexts and subsequently the estimate for the across-individual factor decrease in the odds of buzzing during *LFAS* was high (0.25 [0.21, 0.84]). Some 30% of the variability was unexplained by the observed changes in the state budget. These results demonstrate how variability in buzz rates can make both the detection and interpretation of an anthropogenic effect on prey intake challenging. The inclusion of transit vs. layer-restricted search state in models for buzzing remains a robust first step for capturing some of this variability. Chapter 3 illustrated how hidden state processes can incorporate empirically derived relationship between depth and vertical steps. Hidden state models could also incorporate buzzing, but their interpretation as a part of a functional

('effort') state requires more careful assessment. Chapter 5 explores how sperm whales may adjust their hunting and sensory strategies as a function of foraging depth.

4. References

Allen AN, Schanze JJ, Solow AR, Tyack PL (2014) Analysis of a Blainville's beaked whale's movement response to playback of killer whale vocalizations. Mar Mammal Sci 30:154–168

Alves A, Antunes R, Bird A, Tyack PL, Miller PJO, Lam F-P a., Kvadsheim PH (2014) Vocal matching of naval sonar signals by long-finned pilot whales (Globicephala melas). Mar Mammal Sci 30:1248-1257

Antunes R, Kvadsheim PH, Lam FP a, Tyack PL, Thomas L, Wensveen PJ, Miller PJO (2014) High thresholds for avoidance of sonar by free-ranging long-finned pilot whales (Globicephala melas). Mar Pollut Bull

Balcomb K, Claridge D (2001) A mass stranding of cetaceans caused by naval sonar in the Bahamas. Bahamas J Sci 5:2–12

Beale CM (2007) The behavioral ecology of disturbance responses. Int J Comp Psychol 20:111–120

Beale CM, Monaghan P (2004) Behavioural responses to human disturbance: a matter of choice? Anim Behav 68:1065–1069

Bejder L, Samuels a, Whitehead H, Finn H, Allen S (2009) Impact assessment research: use and misuse of habituation, sensitisation and tolerance in describing wildlife responses to anthropogenic stimuli. Mar Ecol Prog Ser 395:177–185

Bruyn PJN de, Tosh C a, Terauds A (2013) Killer whale ecotypes: is there a global model? Biol Rev Camb Philos Soc 88:62–80

Burnham KP, Anderson DR (2002) Model selection and multimodel inference: a practical information-theoretic approach. Springer

Clark C, Ellison W, Southall B, Hatch L, Parijs S Van, Frankel a, Ponirakis D (2009) Acoustic masking in marine ecosystems: intuitions, analysis, and implication. Mar Ecol Prog Ser 395:201–222

Curé C, Antunes R, Alves A (2013) Responses of male sperm whales (Physeter macrocephalus) to killer whale sounds: implications for anti-predator strategies. Sci Rep 3:1579

D'Amico A, Gisiner RC, Ketten DR, Hammock J a., Johnson C, Tyack PL, Mead J (2009) Beaked Whale Strandings and Naval Exercises. Aquat Mamm 35:452–472

Deecke VB, Ford JKB, Slater PJB (2005) The vocal behaviour of mammal-eating killer whales: communicating with costly calls. Anim Behav 69:395–405

Deruiter SL, Southall BL, Calambokidis J, Zimmer WMX, Sadykova D, Falcone E a, Friedlaender AS, Joseph JE, Moretti D, Schorr GS, Thomas L, Tyack PL (2013a) First direct measurements of behavioural responses by Cuvier's beaked whales to mid-frequency active sonar. Biol Lett 9:20130223

DeRuiter SL, Boyd IL, Claridge DE, Clark CW, Gagnon C, Southall BL, Tyack PL (2013b) Delphinid whistle production and call matching during playback of simulated military sonar. Mar Mammal Sci 29:E46–E59

Domenici P, Blagburn JM, Bacon JP (2011) Animal escapology I: theoretical issues and emerging trends in escape trajectories. J Exp Biol 214:2463–73

Dunlop R a, Noad MJ, Cato DH, Kniest E, Miller PJO, Smith JN, Stokes MD (2013) Multivariate analysis of behavioural response experiments in humpback whales (Megaptera novaeangliae). J Exp Biol 216:759–70

Ellison WT, Southall BL, Clark CW, Frankel AS (2011) A new context-based approach to assess marine mammal behavioral responses to anthropogenic sounds. Conserv Biol 26:21–8

Fahlman A, Tyack PL, Miller PJO, Kvadsheim PH (2014) How man-made interference might cause gas bubble emboli in deep diving whales. Front Physiol 5:1–6

Fernández A, Edwards JF, Rodríguez F, Espinosa de los Monteros A, Herráez P, Castro P, Jaber JR, Martín V, Arbelo M (2005) "Gas and fat embolic syndrome" involving a mass stranding of beaked whales (family Ziphiidae) exposed to anthropogenic sonar signals. Vet Pathol 42:446–57

Filadelfo R, Mintz J, Michlovich E, D'Amico A, Tyack PL, Ketten DR (2009) Correlating Military Sonar Use with Beaked Whale Mass Strandings: What Do the Historical Data Show? Aquat Mamm 35:435–444

Finneran JJ, Jenkins AK (2012) Criteria and thresholds for US Navy acoustic and explosive effects analysis. In: Space and naval warfare systems center pacific San Diego.

Frantzis a (1998) Does acoustic testing strand whales? Nature 392:29

Frid A, Dill L (2002) Human-caused disturbance stimuli as a form of predation risk. Conserv Ecol 6

Gill J, Norris K, Sutherland W (2001) Why behavioural responses may not reflect the population consequences of human disturbance. Biol Conserv 97:265–268

Goldbogen J a, Southall BL, Deruiter SL, Calambokidis J, Friedlaender AS, Hazen EL, Falcone E a, Schorr GS, Douglas A, Moretti DJ, Kyburg C, McKenna MF, Tyack PL (2013) Blue whales respond to simulated mid-frequency military sonar. Proc Biol Sci 280:20130657

Hildebrand J (2009) Anthropogenic and natural sources of ambient noise in the ocean. Mar Ecol Prog Ser 395:5–20

Hooker SK, Baird RW, Fahlman A (2009) Could beaked whales get the bends? Effect of diving behaviour and physiology on modelled gas exchange for three species: Ziphius cavirostris, Mesoplodon densirostris and Hyperoodon ampullatus. Respir Physiol Neurobiol 167:235–46

Hotchkin C, Parks S (2013) The Lombard effect and other noise-induced vocal modifications: insight from mammalian communication systems. Biol Rev Camb Philos Soc

Johnson M, Aguilar de Soto N, Madsen PT (2009) Studying the behaviour and sensory ecology of marine mammals using acoustic recording tags: a review. Mar Ecol Prog Ser 395:55–73

Kvadsheim PH, Miller PJO, Tyack PL, Sivle LD, Lam FP a, Fahlman a (2012) Estimated Tissue and Blood N(2) Levels and Risk of Decompression Sickness in Deep-, Intermediate-, and Shallow-Diving Toothed Whales during Exposure to Naval Sonar. Front Physiol 3:125

Lusseau D, Bejder L (2007) The long-term consequences of short-term responses to disturbance experiences from whalewatching impact assessment. Int J Comp Psychol 20:228–236

Madsen PT (2005) Marine mammals and noise: Problems with root mean square sound pressure levels for transients. J Acoust Soc Am 117:3952

Madsen PT, Payne R, Kristiansen NU, Wahlberg M, Kerr I, Møhl B (2002) Sperm whale sound production studied with ultrasound time/depth-recording tags. J Exp Biol 205:1899–906

Martin A, Clarke M (1986) The diet of sperm whales (Physeter macrocephalus) captured between Iceland and Greenland. Mar Biol Assoc United Kingdom:779–790

McGregor PK (1992) Design of playback experiments. In: Playback and Studies of Animal Communication. Plenum Press, New York

Miller P, Antunes R, Alves A (2011) 3S experiments: studying the behavioural effects of naval sonar on killer whales (Orcinus orca), sperm whales (Physeter macrocephalus), and long-finned. Scottish Ocean Inst Tech Rep

Miller PJO, Antunes RN, Wensveen PJ, Samarra FIP, Catarina Alves A, Tyack PL, Kvadsheim PH, Kleivane L, Lam F-P a., Ainslie M a., Thomas L (2014) Dose-response relationships for the onset of avoidance of sonar by free-ranging killer whales. J Acoust Soc Am 135:975–993

Miller PJO, Johnson MP, Madsen PT, Biassoni N, Quero M, Tyack PL (2009) Using at-sea experiments to study the effects of airguns on the foraging behavior of sperm whales in the Gulf of Mexico. Deep Sea Res Part I Oceanogr Res Pap 56:1168–1181

Miller PJ, Kvadsheim PH, Lam, Frans-Peter A Wensveen PJ, Antunes R, Alves AC, Visser F, Kleivane L, Tyack PL, Sivle LD (2012) The Severity of Behavioral Changes Observed During Experimental Exposures of Killer (Orcinus orca), Long-Finned Pilot (Globicephala melas), and Sperm (Physeter macrocephalus) Whales to Naval Sonar. Aquat Mamm 38:362–401

Møhl B, Wahlberg M, Madsen PT, Heerfordt A, Lund A (2003) The monopulsed nature of sperm whale clicks. J Acoust Soc Am 114:1143

Mooney TA, Hanlon R, Madsen PT, Christensen-dalsgaard J, Ketten DR, Nachtigall PE (2012) Potential for Sound Sensitivity in Cephalopods. In: The Effects of Noise on Aquatic Life (AN Popper and A Hawkins, Eds.). Springer New York, New York, NY

Moore MJ, Early G a (2004) Cumulative sperm whale bone damage and the bends. Science 306:2215

Nowacek DP, Thorne LH, Johnston DW, Tyack PL (2007) Responses of cetaceans to anthropogenic noise. Mamm Rev 37:81–115

Oliveira CC, Wahlberg M, Johnson M, Miller PJO, Madsen PT (2013) The function of male sperm whale slow clicks in a high latitude habitat: communication, echolocation, or prey debilitation? J Acoust Soc Am 133:3135–44

Popper a N, Hastings MC (2009) The effects of anthropogenic sources of sound on fishes. J Fish Biol 75:455–89

Quirós YB de, González-Diaz O, Arbelo M, Sierra E, Sacchini S, Fernández A (2012) Decompression vs. Decomposition: Distribution, Amount, and Gas Composition of Bubbles in Stranded Marine Mammals. Front Physiol 3:177

Reeves R (2003) Dolphins, whales and porpoises: 2002-2010 conservation action plan for the world's cetaceans.

Ridgway S, Carder D (2001) Assessing hearing and sound production in cetaceans not available for behavioral audiograms: Experiences with sperm, pygmy sperm, and gray whales. Aquat Mamm

Santos M, Pierce G, Boyle P (1999) Stomach contents of sperm whales Physeter macrocephalus stranded in the North Sea 1990-1996. Mar Ecol Prog Ser 183:281–294

Schlundt CE, Finneran JJ, Carder D a, Ridgway SH (2000) Temporary shift in masked hearing thresholds of bottlenose dolphins, Tursiops truncatus, and white whales, Delphinapterus leucas, after exposure to intense tones. J Acoust Soc Am 107:3496–508

Sih A (2013) Understanding variation in behavioural responses to human-induced rapid environmental change: a conceptual overview. Anim Behav 85:1077–1088

Simon MJ, Kristensen TK, Kinze CC, Tougaard S, Tendal OS (2003) Gonatus fabricii (Mollusca, Theuthida) as an important food source for sperm whales (Physeter macrocephalus) in the Northeast Atlantic. Sarsia North Atl Mar Sci 88:244–246

Sivle LD, Kvadsheim PH, Fahlman a, Lam FP a, Tyack PL, Miller PJO (2012) Changes in dive behavior during naval sonar exposure in killer whales, long-finned pilot whales, and sperm whales. Front Physiol 3:400

Solé M, Lenoir M, Durfort M, López-Bejar M, Lombarte A, Schaar M van der, André M (2013) Does exposure to noise from human activities compromise sensory information from cephalopod statocysts? Deep Sea Res Part II Top Stud Oceanogr 95:160–181

Southall BLB, Bowles AE, Ellison WT (2007) Marine mammal noise-exposure criteria: initial scientific recommendations. Bioacoustics 33

Stankowich T, Blumstein DT (2005) Fear in animals: a meta-analysis and review of risk assessment. Proc Biol Sci 272:2627–34

Tyack PL (2009) Acoustic playback experiments to study behavioral responses of free-ranging marine animals to anthropogenic sound. Mar Ecol Prog Ser 395:187–200

Tyack PL, Zimmer WMX, Moretti D, Southall BL, Claridge DE, Durban JW, Clark CW, D'Amico A, DiMarzio N, Jarvis S, McCarthy E, Morrissey R, Ward J, Boyd IL (2011) Beaked whales respond to simulated and actual navy sonar. PLoS One 6:e17009

Watts AJ (2003) Underwater Warfare Systems, 2002-2003. Jane's Information Group Limited

Watwood SL, Miller PJ, Johnson M, Madsen PT, Tyack PL (2006) Deep-diving foraging behaviour of sperm whales (Physeter macrocephalus). J Anim Ecol 75:814–825

Williams R, Lusseau D, Hammond PS (2006) Estimating relative energetic costs of human disturbance to killer whales (Orcinus orca). Biol Conserv 133:301–311

Wilson M, Hanlon RT, Tyack PL, Madsen PT (2007) Intense ultrasonic clicks from echolocating toothed whales do not elicit anti-predator responses or debilitate the squid Loligo pealeii. Biol Lett 3:225–7

Wright AJ, Deak T, Parsons ECM (2011) Size matters: management of stress responses and chronic stress in beaked whales and other marine mammals may require larger exclusion zones. Mar Pollut Bull 63:5–9

Zimmer WMX, Tyack PL (2007) Repetitive Shallow Dives Pose Decompression Risk in Deep-Diving Beaked Whales. Mar Mammal Sci 23:888–925

4. Tables & Figures

Table 4.1. Sample size in each exposure condition (h)

	Total	Baseline	Sonars						Orca playbacks			
			Control		MFAS		LFAS		Control		Signal	
Tag id			Dur	Post	Dur	Post	Dur	Post	Dur	Post	Dur	Post
sw05_196a	20.8 (14)	20.8 (14)										
sw05_199a	17.6 (14.8)				4.3 * (3.7)	13.3 * (11.1)						
sw05_199b	13.3 (9.7)				3.1 * (2.5)	10.2 * (7.2)						
sw05_199c	12.9 (6.5)				3.6 * (2.8)	9.3 * (3.6)						
sw08_152a	5.8 (4.4)	1.8 (1.4)			1.6 (1.4)	1 (0.8)	1 (0.4)	0.4 (0.3)				
sw09_141a	14.3 (5.6)	3 (1)	1.1 (0.8)	2.1 (0.9)	0.9 (0.8)	1 (0.7)	0.7 (0.1)	1 (0.3)			0.4 (0.2)	4.2 (0.7)
sw09_142a	14.5 (7.7)	2.8 (2.4)	0.6 (0.5)	1 (0.4)	0.6 (0.5)	2.5 (0.3)	1.6** (0.4)	4** (2.8)			0.3 (0)	1.1 (0.6)
sw09_153a	8.4 (7.2)	8.4 (7.2)										
sw09_160a	14.5 (12.1)	3.2 (2.8)			0.7 (0.6)	1.7 (1.4)	1.7** (1.3)	5.3** (4.5)			0.3 (0.3)	1.5 (1.2)
sw10_147a	14.8 (8.1)	14.8 (8.1)										
sw10_149a	14.3 (11.2)	10.1 (8.1)							0.3 (0.3)	0.4 (0.2)	0.3 (0)	1 (0.7)
sw10_150a	14.1 (10.8)	8.8 (7)							0.3 (0.3)	0.2 (0.1)	0.3 (0.2)	1.2 (0.8)

* incidental MFA exposures; ** Includes both LFA up-sweep and down-sweep exposures.

Brackets: time spent in foraging states (descent, layer-restricted search and ascent), Dur: during exposure duration, Post: post-exposure durations, refer to time period after exposure to next exposure, or end of tag record.

Table 4.2. Candidate exposure covariates

Covariate type	Explanation	Abbreviations Presence/ absence	Presence/absence + n min window	Time decay
Incidental sonar				
	Presence/absence of received pings in the last n min		*SON05_10* *SON05_15* *SON05_30*	*minFromSON05*
	Maximum $SPL_{90\%}$ in the last n min		*SPL_max_10* *SPL_max_15* *SPL_max_30*	
Experiment				
	Approach of the sonar source vessel	*BoatPass*	*BoatPass_win*	*minFromBoatPass* *min²FromBoatPass*
	Order of sonar source vessel approach	*BoatPass #*		
	Orca playback boat in operation (signal or noise playback)	*OrcaPB*	*OrcaPB_win*	
Signal				
	Low-frequency active sonar playback	*LFAS*	*LFAS_win*	*minFromLFAS* *min²FromLFAS*
	Medium-frequency active sonar playback	*MFAS*	*MFAS_win*	*minFromMFAS* *min²FromMFAS*
	Orca signal playback	*OrcaS*	*OrcaS*	*minFromOrcaS* *min²FromOrcaS*
	Experimental sonar source level	*SL*		
	Experimental sonar received sound pressure level (SPL_{max})	*RL*		
	Experimental sonar cumulative exposure level	*SELcum*		

Presence/absence covariate values were set 1 during exposure and 0 otherwise. Presence/absence + time window covariate values were set 1 during exposure, 1 during win min into post-exposure, and 0 otherwise. Time-decay covariates were calculated as time (min) or time squared (min^2) respectively, since last exposure, set to zero during exposure and 900 or 900^2 min for baseline.

Table 4.3. Binomial GEE model for buzzing: incidental exposures

Parameter	Estimate	SE	95% CI		z	p
Intercept	-2.56	0.26	-3.06	-2.05	-9.91	0.000
Descent	0.96	0.23	0.51	1.41	4.19	0.000
LRS	1.36	0.24	0.89	1.84	5.63	0.000
SON05_10	0.55	0.29	-0.02	1.12	1.90	0.057

Parameter estimates, empirical standard errors and Wald test statistics are shown. The model was specified whale as a cluster variable and independence as a working correlation.

Table 4.4 Pre-exposure baseline vs. experimental exposure time budgets and energetic proxies

Tag id	Exposure	duration	Time spent (% bins)						Buzz presence (%)		mean ODBA	
		(min)	surface	descent	LRS	ascent	resting	silent. a	obs.	expected	obs.	expected
sw08_152a	baseline	106	18.9	12.3	63.2	5.7	0.0	0.0	12.3	12.3	25.1	25.1
	MFAS	96	12.5	26.0	37.5	22.9	0.0	1.0	9.4	7.3	22.9	24.3
	LFAS	61	8.2	13.1	3.3	23.0	0.0	52.5	0.0	0.6	23.4	23.7
sw09_141a	baseline	181	18.2	4.4	24.9	5.0	1.1	46.4	3.9	3.9	25.7	25.7
	LFAS	41	2.4	0.0	2.4	9.8	0.0	85.4	2.4	1.4	34.6	24.9
	MFAS	53	9.4	15.1	62.3	9.4	0.0	3.8	9.4	9.9	26.1	26.5
	Sonar ctrl	65	16.9	10.8	56.9	7.7	0.0	7.7	3.1	8.5	24.3	26.4
	OrcaS	22	4.5	18.2	22.7	13.6	0.0	40.9	4.5	6.3	25.1	26.3
sw09_142a	baseline	165	12.7	6.7	73.9	5.5	0.0	1.2	8.5	8.5	24.8	24.8
	Sonar ctrl	36	13.9	11.1	66.7	8.3	0.0	0.0	2.8	8.0	22.1	24.5
	LFAS	45	20.0	8.9	0.0	15.6	0.0	55.6	0.0	1.7	22.2	23.7
	MFAS	34	8.8	20.6	70.6	0.0	0.0	0.0	20.6	7.5	29.5	25.4
	OrcaS	20	20.0	0.0	0.0	0.0	20.0	60.0	0.0	0.0	20.5	20.9
	LFAS-DS	51	17.6	11.8	0.0	7.8	5.9	56.9	0.0	0.9	17.8	23.2
sw09_160a	baseline	195	12.8	18.5	54.4	14.4	0.0	0.0	36.4	36.4	25.6	25.6
	MFAS	43	23.3	9.3	39.5	27.9	0.0	0.0	16.3	27.6	24.7	24.5
	LFAS	44	15.9	6.8	40.9	11.4	0.0	25.0	0.0	26.1	21.3	25.6
	OrcaS	20	0.0	40.0	60.0	0.0	0.0	0.0	20.0	42.3	22.3	25.5
	LFAS-DS	60	10.0	30.0	45.0	11.7	0.0	3.3	13.3	33.0	22.3	24.6
sw10_149a	baseline	603	17.7	15.3	52.2	12.9	0.2	1.7	9.8	9.8	24.8	24.8
	Orca ctrl	16	0.0	43.8	56.3	0.0	0.0	0.0	6.3	10.3	22.8	25.5
	OrcaS	16	6.3	6.3	0.0	6.3	0.0	81.3	0.0	0.9	26.3	24.4
sw10_150a	baseline	531	20.2	23.0	37.5	18.6	0.0	0.8	16.9	16.9	28.2	28.2
	Orca ctrl	16	0.0	43.8	56.3	0.0	0.0	0.0	18.8	26.8	33.5	33.9
	OrcaS	16	37.5	31.3	12.5	18.8	0.0	0.0	0.0	11.6	24.8	24.4

Table 4.5. GLM and GEE model selection for experimental exposures

Response	Covariate	AIC selection			GEE	Wald (full)		Wald (final)	
		Df	AIC	ΔAIC	Df	χ^2	p	χ^2	p
state	full model		7468.4	126.1					
S. active	*prevState*	25	23813.0	16344.6	5	542.1	0.000	527.7	0.000
Foraging					5	2962.2	0.000	2829.4	0.000
	whale	55	7665.3	197.0					
S. active	*LFAS_8*	5	7558.1	89.8	1	35.6	0.000	30.5	0.000
Foraging					1	19.5	0.000	19.2	0.000
S. active	*OrcaS_19*	5	7528.8	60.5	1	82.9	0.000	51.3	0.000
Foraging					1	6.2	0.013	7.7	0.006
S. active	*minFromOrcaS2*	5	7495.0	26.6	1	2.9	0.087		
Foraging					1	0.3	0.561		
buzz	full model		5058.6	33.9					
	state	2	5378.5	319.9	2	96.7	0.000	106.0	0.000
	whale	11	5380.7	322.1					
	BoatPass	1	5088.4	29.7	1	59.8	0.000	26.2	0.000
	MFAS	1	5079.6	21.0	1	1250.9	0.000	11.6	0.001
	OrcaPB	1	5061.0	2.3	1	12.5	0.000	7.9	0.005
	minFromMFAS	1	5065.6	7.0	1	0.6	0.447		
	minFromBoatPass2	1	5061.5	2.8	1	0.7	0.413		
ODBA	full model		67969.7	31.4					
	state	5	70613.2	2643.4	5	632.4	0.000	622.8	0.000
	whale	11	68286.2	316.4					
	noBoatPass	1	68001.2	31.5	1	9.3	0.002		
	MFAS_13	1	67985.4	15.6	1	2.1	0.151		
	LFAS	1	67990.8	21.0	1	9.8	0.002		
	minFromLFAS	1	67971.5	1.8	1	4.4	0.036	3.9	0.049

Full model: AIC selected model on the left, with AIC and ΔAIC to the baseline model shaded for each response variable (*state*, *buzz*, *ODBA*). AIC and ΔAIC for each covariate show change in AIC when dropped from the AIC selected effects model. Wald (full) test statistics and p-values are shown for both the full (AIC selected) and reduced (final model in backwards selection) GEE models. Df refers to degrees of freedom of each covariate.

Table 4.6. Binomial GEE model for silent active and foraging states: experimental exposures

	Silent active state			Foraging state		
Parameter	Estimate	95% CI		Estimate	95% CI	
Intercept s. active	0.27	-0.84	1.37	-1.50	-2.17	-0.83
prevState surface	-3.36	-4.27	-2.46	-0.92	-1.62	-0.23
prevState descent	-4.86	-5.57	-4.14	5.73	5.19	6.26
prevState LRS	-5.79	-6.34	-5.23	6.42	6.08	6.75
prevState ascent	-5.78	-7.02	-4.53	2.98	2.28	3.69
prevState drifting	-3.23	-4.34	-2.11	-2.77	-4.12	-1.41
LFAS_8	1.87	1.21	2.54	-0.90	-1.31	-0.50
OrcaS_19	1.75	1.27	2.22	-0.93	-1.58	-0.27

Parameter estimates and 95% confidence intervals.

Table 4.7. Binomial GEE model for probability of buzzing: experimental exposures

Parameter	Estimate	SE	95% CI		z	p
Intercept	-2.86	0.22	-3.29	-2.42	-12.86	<0.001
Descent	1.02	0.28	0.48	1.57	3.69	<0.001
LRS	1.72	0.20	1.32	2.12	8.45	<0.001
BoatPass	-1.27	0.25	-1.76	-0.78	-5.12	<0.001
MFAS	0.99	0.29	0.42	1.56	3.41	0.001
OrcaPB	-0.51	0.18	-0.87	-0.15	-2.81	0.005

Parameter estimates, empirical standard errors and Wald test statistics are shown. The model was specified whale as a cluster variable and independence as a working correlation.

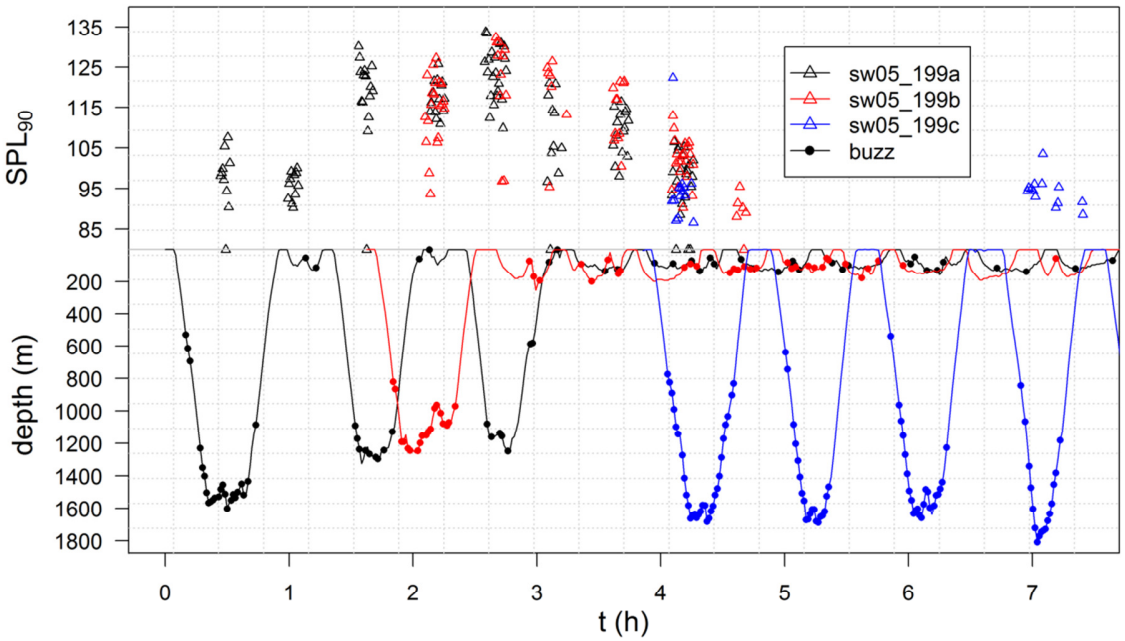


Figure 4.1. Incidental sonar detections and received levels

Received sonar level (triangles, $SPL_{90\%}$ dB rms re 1µPa) of incidental sonar at each tag overlaid with dive profile (1-minute down-sampled data) and presence/absence of buzzes (solid circles). Detected pings that could not be estimated $SPL_{90\%}$ are shown at grey horizontal line (depth=0).

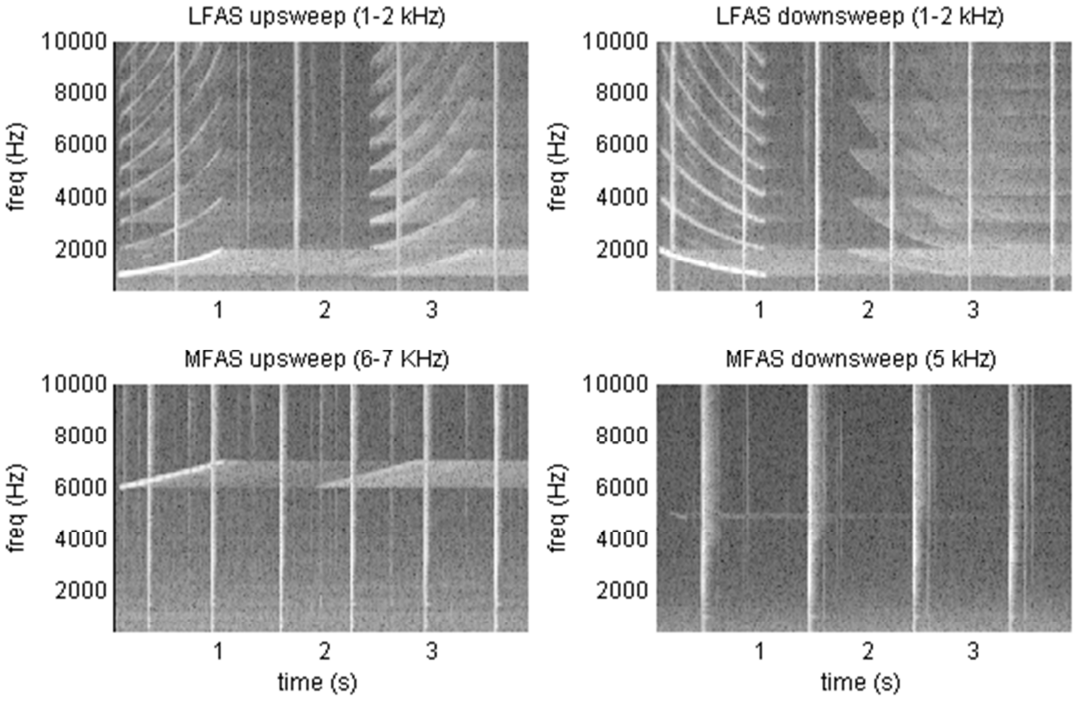


Figure 4.2. a Example spectral signatures of each sonar type

Experimental sonar pings (LFAS upsweep, LFAS downsweep and MFAS upsweep) were 1 s in duration, while incidental MFAS downsweep signal (bottom right) duration was 0.15 s. Harmonic distortions for LFAS are shown (top). Broadband transients are tagged whale regular clicks. Experimental sonars examples were extracted from sw09_160a at full power phase, and incidental sonar was extracted from sw05_199a at 2 h 18 min (Figure 4.1).

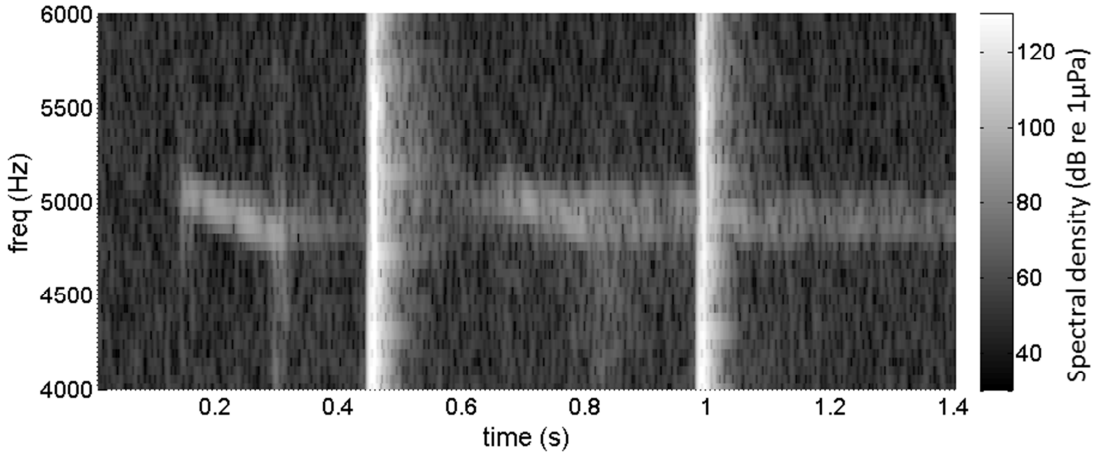


Figure 4.2. b Example incidental MFAS downsweep

A close-up of the incidental MFAS downsweep in Figure 4.2 a. Signal downsweep is shown at 0.15-0.3 s with reverberation at 0.6-1.4s and two tagged whale clicks at 0.45s and 1s (Blackman harris window with 50% overlap, resolution 2048 samples at 96 kHz).

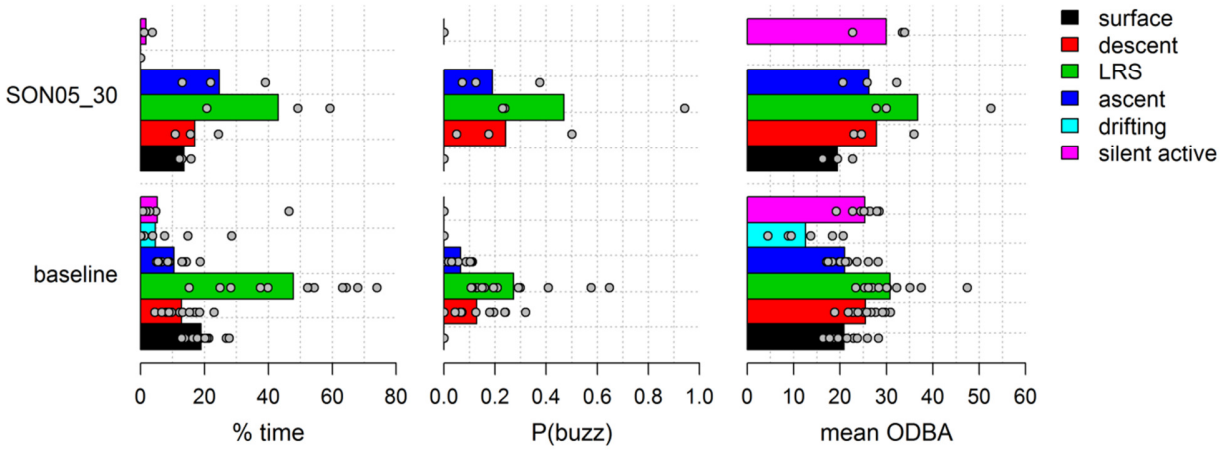


Figure 4.3. Time and energetic budgets during baseline vs. incidental sonar exposures

Left: behavioural time budgets, middle: proxies of foraging success (probability of buzzing, % of bins) and right: locomotion cost (mean ODBA) averaged across each individual for each state during incidental sonar (SON05_30, 8.7 h) and baseline (108.8 h, post-exposure of incidentally exposed whales, and pre-exposure baseline from all other whales). Solid circles show within-individual time budgets.

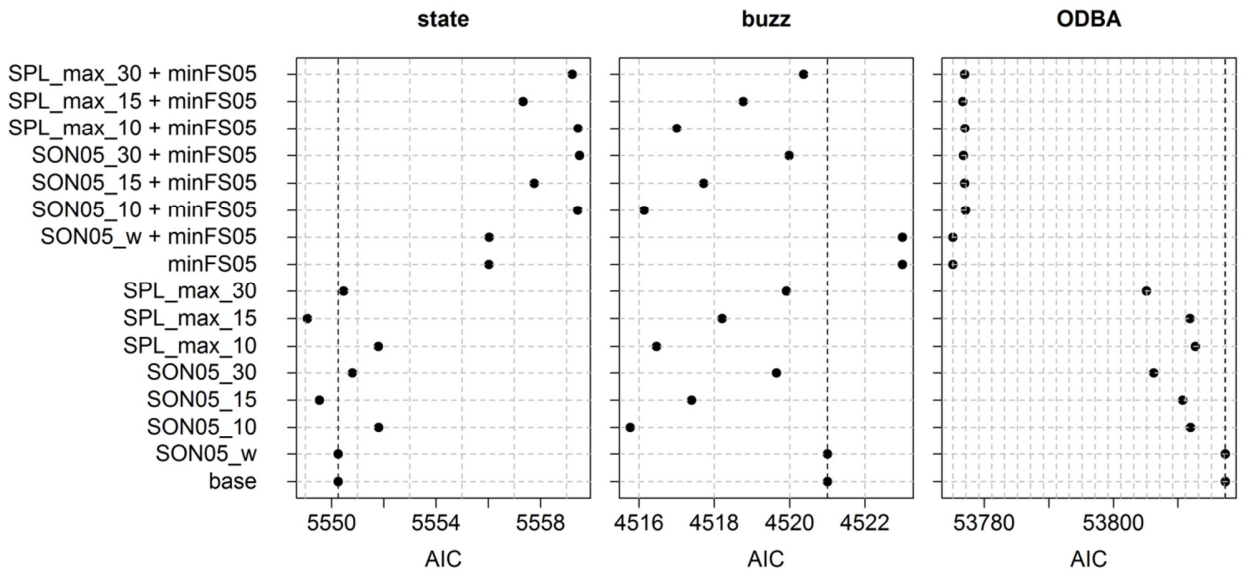


Figure 4.4. AIC model selection for incidental sonar exposures

Horizontal lines show AIC differences of 2 units. Black dashed line show the AIC of the base models (state ~ prevState + whale, buzz ~ state + whale, ODBA ~ state + whale).

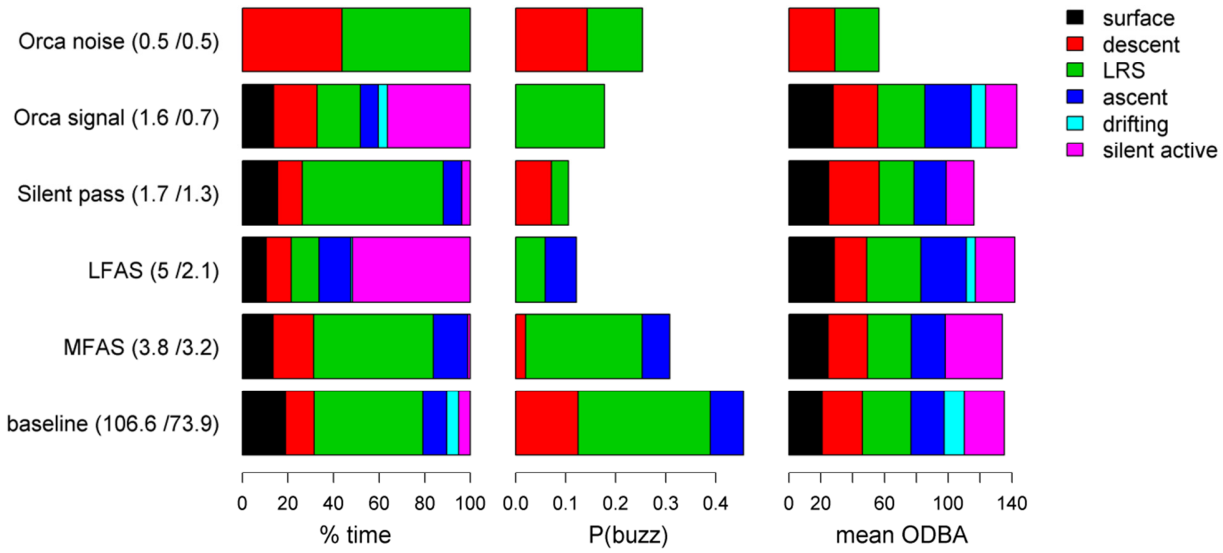


Figure 4.5. Time and energetic budgets during baseline vs. experimental exposures

Left: behavioural time budgets, middle: proxies of foraging success (probability of buzzing, % of bins) and right: locomotion cost (mean ODBA) averaged across each individual for each state during each experimental exposure and pre-exposure baseline (excluding tagging periods and incidental sonar periods SON05_30). Sample sizes (h) are given both including and excluding non-foraging states (surfacing, drifting and silent active).

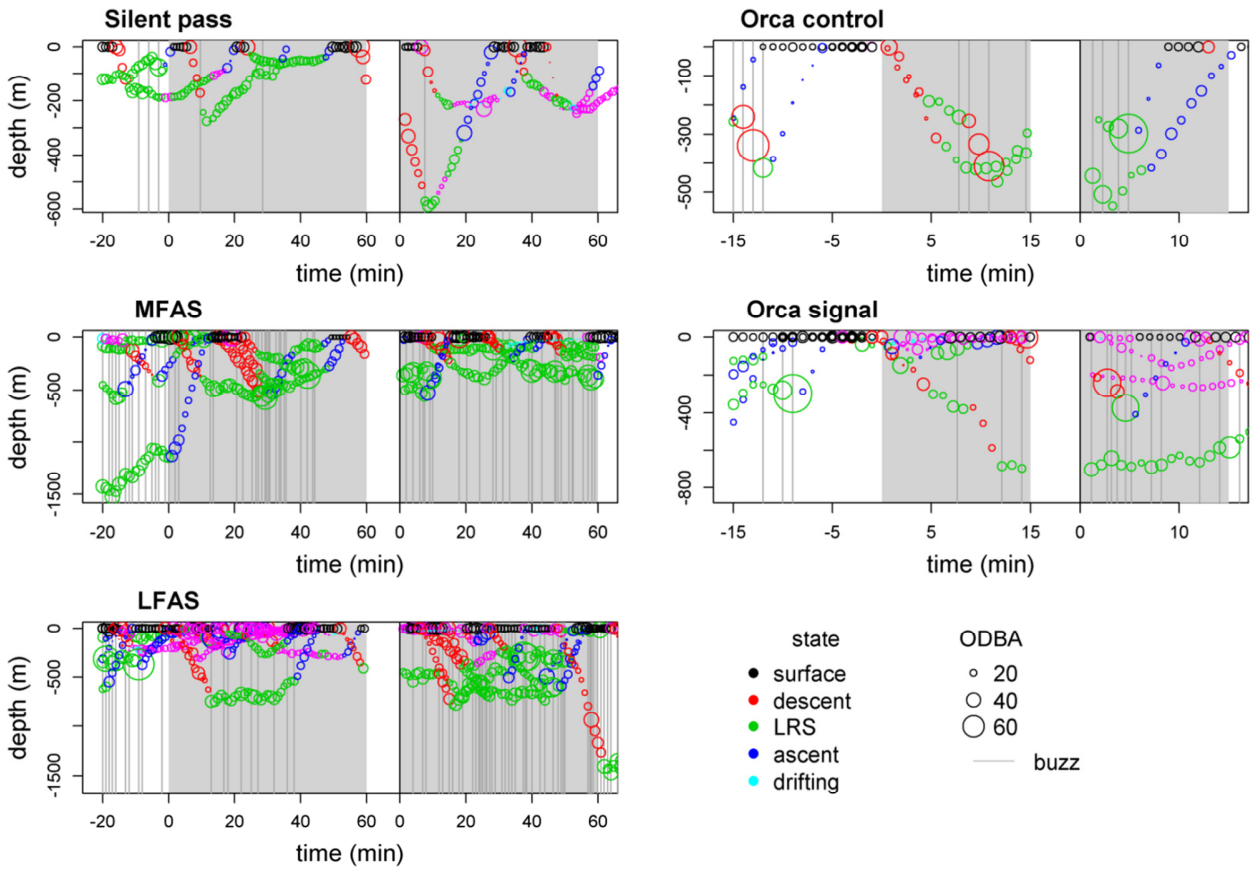


Figure 4.6. Response profiles for each experiment

Shaded areas on the left show data from during experiment, and shaded areas on the right show data from after experiment. Times are given from start and end of experiment, respectively. Vertical grey lines show buzz presence and circle size shows 2-norm ODBA summed over each 1-minute time interval.

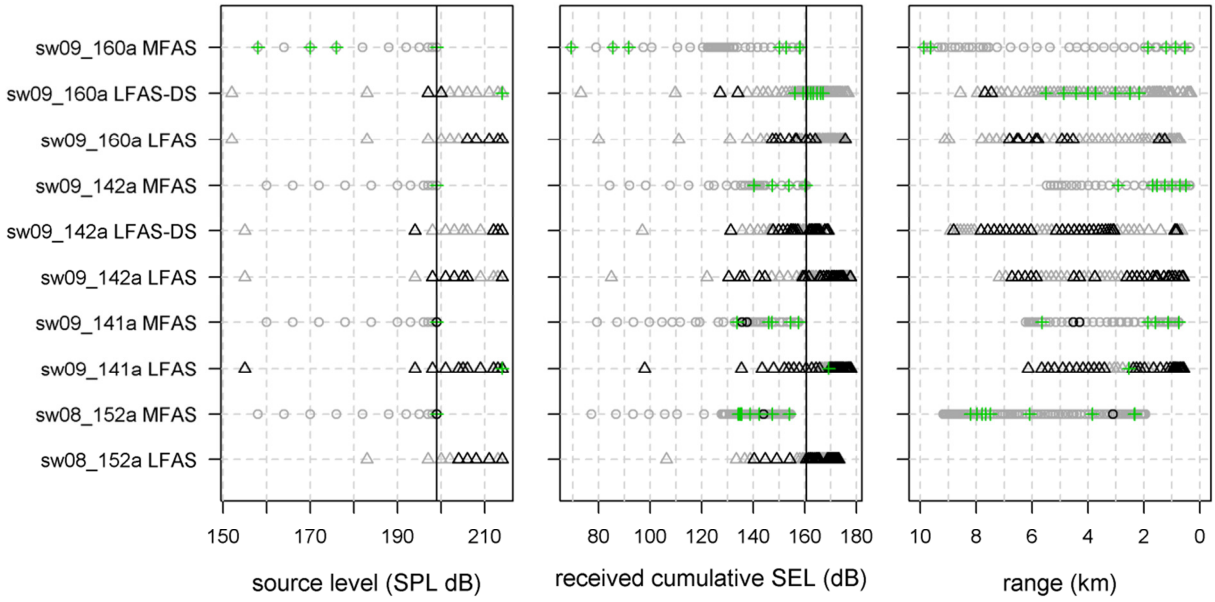


Figure 4.7 Source level, received level and range contrast between LFAS and MFAS exposures

Black: Silent active states and green: presence/absence of buzzes as a function of sound pressure level at source (SL), cumulative sound exposure level (SEL) and estimated range between the source and whale. Circles: data from MFAS experiments, triangles: data from LFAS experiments.

5 SPERM WHALE MOVEMENT AND ACOUSTIC SAMPLING STRATEGIES DURING PREY ENCOUNTERS: AN EASIER MEAL AT DEPTH?

5.1 Abstract

The extreme diving capabilities of some marine mammals highlight an adaptive benefit to foraging at great depths. However, relatively little is known about the foraging strategies that deep-diving cetaceans employ to target prey at different depths, or how they find and select prey. In this chapter, tag-recorded movement and acoustic metrics during buzz production by the tagged whale were used as indicators of foraging behaviour that might vary with different types of prey. Movement and acoustic metrics were analysed separately in order to decompose physical effects of depth (pressure, light) from the predator's movement strategy during prey capture which is indicated by buzzing. Movement metrics were aimed to reflect variations in prey manoeuvrability and responsiveness by quantifying the predator's movement path (vertical velocity, pitch and heading variance) and locomotion effort (ODBA, Chapters 2-3). The spatial density of selected prey was quantified as time between prey capture attempts (inter-buzz-interval IBI). A cluster analysis of the IBI and the movement metrics revealed a highly depth-dependent distribution of four clusters, here descriptively labelled as "sparse-shallow", "mid-active", "deep-descent" and "deep" buzz types. Multivariate regression modelling of acoustic metrics with buzz type and depth as explanatory variables revealed that maximum click rate in a buzz, but not apparent click output level (AOL, i.e. off-axis level received by the tag), was significantly related to the buzz type. While click rate was positively correlated with AOL and increased manoeuvring within buzzes, both maximum click rate and AOL had a strong negative and linear relationship with depth across buzzes. These results indicated that while buzz production may have been limited at depth, higher acoustic sampling rates were used to track more mobile prey. Deep-descent and deep buzz types incurred the lowest locomotion effort and were primarily produced during deep dives (median depths 1029 m and 1388, respectively). Their short IBIs, short duration and low click rates suggested that these buzzes were used to hunt less mobile prey, such as immobile female *Gonatus fabricii*. In contrast, sparse-shallow and mid-active buzzes (median depths 265 m and 404 m, respectively) had a more variable IBI, had higher buzz click rates, were often longer in duration and coincided with more

energetic manoeuvres. This indicated a more patchy distribution (or a higher selection threshold) of difficult to capture and potentially more energy-rich pelagic prey, such as muscular cephalopods or fish. Given a deep-diving predator must balance both diving and energetic costs of foraging at depth, foraging on less mobile and perhaps more abundant prey likely compensated for higher transport costs at great depths, even if echolocation rates were limited due to pressure effects on the sound production mechanism.

5.2 Background

Chapters 2-4 used presence of buzzing as a proxy for foraging success in sperm whales. Dive-scale patterns in movement and clicking were deemed to represent foraging effort. At a finer temporal scale however, buzzing also represents "effort" because the actual rate of prey capture is unknown. Indeed, relatively little is known about how deep-diving marine mammals select and capture their prey, and what might be the different benefits and costs of foraging upon different prey types at different depth layers. Cephalopods are an important source of biomass in marine food webs, and their availability at depth may have selected for a larger body size and more extreme diving capabilities in a range of squid-eating marine mammals (Clarke 1996, Slater et al 2010). Deep-diving marine mammals whose diets overlap with the sperm whale include elephant seals, certain species of beaked whales (e.g. *Hyperoodon*), pygmy and dwarf sperm whales (Clarke 1996, Whitehead et al 2003). While odontocetes use echolocation as a means to find prey, seals apparently lack this ability and may use vision and tactile senses instead (e.g. hydrodynamic trail following; Dehnhardt et al 2001). Echolocation may allow odontocetes a more selective feeding strategy than seals (Naito et al 2013). Nevertheless, one advantage of studying echolocating species is that the portion of the environmental information that is available through echolocation be quantified (e.g. Wisniewska et al 2012) while insights about their ecological niche can be relevant across deep-diving marine mammals that play key roles as top predators in marine food webs (Heithaus et al 2008) and flagship species for marine conservation (Sergio et al 2008).

The main component of sperm whale diet, cephalopods, provides a diverse food source ranging from muscular and protein rich cephalopods to neutrally buoyant ammoniac squids that are lower in energy content (Kawakami 1980, Clarke 1996, Santos et al 1999, Santos et al 2002). Sperm whales can also take fish, which may be particularly important in the high-latitude foraging grounds in New Zealand (Gaskin and Cawthorn 1967) and North Atlantic (Martin and Clarke 1986). Many species of cephalopod may be relatively easy to catch as they quickly become exhausted after fast swimming (Clarke et al 1996) and they may be targeted at more vulnerable stages

of vertical migration (e.g. jumbo squid in Gulf of California, Davis et al 2007) or ontogenetic stage (Simon et al 2003). In the Norwegian sea for example, sperm whale feeding grounds overlap with the terminal foraging grounds of *Gonatus fabricii* (Bjørke 2001). This species has been described as the most prevalent prey type in the stomachs of stranded specimens in the North Sea and Norwegian coast (Santos et al 2002, Simon et al 2003). It has been suggested that sperm whales and pilot whales (*Globicephalas melas*) exploit aggregations of cephalopods that are either dead or are spent after spawning (Clarke 1996), and that sperm whales may also target egg carrying female *Gonatus fabricii* (Simon et al 2003). During the egg carrying phase, which extend through for the year before hatching, the gravid females are effectively immobilised (Bjørke 2001).

While the animal-attached DTAG has enabled measurements associated with likely prey encounters (Miller et al 2004, Johnson et al 2004, Madsen et al 2013), uncertainties remain about the type of prey that is targeted at different depths (Teloni et al 2008, Aoki et al 2012). Understanding variability in prey capture rates and subsequent energetic consequences is particularly important if buzz rates are being used as a proxy for feeding rate during anthropogenic effects studies (Chapters 2-4). By recording of buzz duration and inter-buzz-intervals, Teloni et al (2008) showed that sperm whales produced more frequent and shorter buzzes during deep (>800 m) compared to shallow dives, indicating more sedentary and densely distributed prey at greater depths. Similarly, Aoki et al (2012) observed two types of speed bursts at depths of 400-800 m, “inflectional” and “linear”: inflectional bursts involved more rapid changes in orientation and multiple acceleration phases, while linear bursts typically consisted of a single acceleration event preceded by gliding and followed by a gradual deceleration. Furthermore, their study recorded bursts of speed during bottom-phases of dives, but at rates that appeared inconsistent for all prey captures to require an active burst (Aoki et al 2012). By integrating of both movement and acoustic data (Miller et al 2004), there is increasing scope to make more accurate inferences about the distribution and manoeuvrability, and subsequently the energetic value, of targeted prey.

In this chapter, tag-recorded movement and acoustic data from aurally verified terminal echolocation buzzes for 12 individuals tagged near Lofoten Islands (Chapters 3-4) were explored. This dataset included the four DTAG records described in Teloni et al (2008). The first objective of this chapter was to examine any clustering of movement during prey encounters whether these clusters ('movement strategies') differ as a function of depth. A vertical distribution of movement might thus indicate if there were differences in prey availability and perhaps preference.

The second objective was to examine whether echolocation behaviour was related to any movement strategies and potentially prey type. It is necessary to consider potential effects of pressure on sound production, as both pressure and prey availability are likely to differ by depth. Sperm whale sound production is thought to be pneumatically driven (Ridgway and Carder 1988, Huggenberger et al 2012), and the limited volume of gas available to a breath holding deep diver suggests that gas must be recycled. Several observations of pauses in between bouts of regular clicks and following buzzes (e.g. Wahlberg 2002), and air-movement sounds that can be heard during these pauses on DTAG recordings (Pers. Obs.) strongly suggest that pauses function as air-recycling events. Thus, changes in gas volume or density may influence aspects of echolocation output. The number of regular clicks produced in between recycling events (i.e. regular click train duration) has been shown to decrease with depth, consistent with gas volumes being reduced by hydrostatic pressure (Wahlberg 2002). However, sperm whale regular click levels and frequency content appear to be relatively unaffected by pressure (Madsen et al 2002). Nevertheless, if a certain gas volume is required to produce each click, reduced gas volume under pressure could potentially limit buzz production more than regular clicks.

Echolocation requires a tight coupling of both sensory and motor systems, and echolocation tactics can vary depending on a behavioural mode (e.g. between searching and pursuit of prey) and with environmental parameters (e.g. clutter and ambient noise conditions) (Schnitzler et al 2003, Wisniewska et al 2012, Madsen and Surlykke 2013). Relatively fast click rates, or inversely, short inter-click-intervals (ICI), coupled with lower source levels during terminal echolocation appear to be a common

feature in all studied odontocetes, and indeed most echolocating bat species, highlighting a key function in acoustic gaze control (Madsen and Surlykke 2013). For example, while beaked whale search clicks are frequency-modulated, their broadband buzz clicks resemble those of large delphinids (Johnson et al 2006). Reduced source level and increased sampling rates effectively increase the temporal resolution (data rate) to track likely evasive prey while reducing the complexity of the auditory scene and echo ambiguity (Madsen and Surlukke 2013). Wisniewska et al (2012) suggest that during buzzing, porpoises reduce their depth of gaze to a single target while engaging in a more reactive mode of sensory-motor operation, i.e. sensory volume is reduced relative to the motor (or stopping) volume (Wisniewska et al 2012).

Echolocation output may also be partially adjusted to compensate for decreasing transmission loss (TL) with decreasing range to target, a phenomenon termed automated gain control (AGC, Au 1993). AGC can be achieved by adjustments to source level, hearing sensitivity or both (transmitter vs. receiver-based AGC; Finneran et al 2013, Supin and Nachtigall 2013). The transient sonar equation describes the relationship between the transmitted and received levels of sonar (Urick 1983, Au 1993). Assuming spherical transmission loss (TL) and frequency-dependent absorption α, the equation can be written as:

$$RL = SL - 40\log_{10}(R) - \alpha + TS \qquad Equation\ 5.1$$

Where RL and SL are the received and source energy flux density respectively, R is range and TS is target strength in decibels. Beaked whales (*Mesoplodon*) appear to maintain a relatively constant output on approach to prey (Madsen et al 2005), while little evidence exists for the presence or lack of AGC in sperm whales. Furthermore, the extent to which sperm whales adjust their click output and rates to their targeted prey during terminal echolocation is not known.

As well as examining movement and echolocation tactics across buzzes produced in their natural context, this chapter aims to elucidate how sperm whales adjust their click rates within buzzes. Three hypotheses were explored: 1) ICI:s track target range that is a correlate of output click level due to AGC, 2) higher sampling rates are

required to increase temporal resolution during increased manoeuvring and turning rates, and 3) click production rates are limited by hydrostatic pressure. Multivariate regression modelling supported all three sources of variation, but also that click rates varied with buzz type that was based upon movement strategies alone.

5.3 Methods

5.3.1 Data collection

DTAG acoustic and movement data were collected aboard 12 tagged sperm whales near Lofoten, Norway in 2005 and 2008-2010 (Section 3.2.1). Data were excluded from Tagging periods (Table 3.1), incidental sonar exposures and experimental exposure periods (Table 4.1).

5.3.2 Acoustic data processing

BUZZ AND CLICK DETECTION

DTAG audio recordings were monitored both aurally and visually using spectrograms in Adobe Audition, and the start and end time of regular and buzz click trains were marked. Buzz start time was defined as a change in amplitude and/or spectral content of clicks before a fast run. Buzz end time was defined as the start of a pause before the next regular click train, exceeding the ICI of the subsequent regular clicks. In the absence of such a clear pause, the end time of a buzz was identified as the last irregularly spaced buzz clicks (this pattern was also typical at the end of the buzzes with a clear pause). For analysis, buzzes were filtered by maximum repetition rate (Section 5.2.3).

Buzz clicks were detected automatically using a custom-built program in Matlab. To enhance signal-to-noise ratio (SNR) for click detection, sound pressures were first band-pass filtered between 700 Hz and 40 kHz using a 256 point finite-impulse filter. Filtered energy was smoothed (Hanning window 1 ms), and click start and end times were detected based upon thresholds of the median and spread of the smoothed energy. See Appendix 5.A for Matlab script.

LEVEL MEASUREMENTS TO CALCULATE 'APPARENT OUTPUT'

The received level of clicks arriving on an animal attached recorder (DTAG) was used as a proxy for the relative acoustic output level or 'apparent output level' (AOL, Madsen et al 2005). DTAG attachment location will vary between individual attachments and potentially also within a tag attachment (Johnson et al 2009) if the tag slides over the animal's body. Therefore, instead of a absolute source levels, the analysis aimed to assess relative changes in AOL within each buzz (Appendix 5C).

In order to measure AOL, a lower order filtering (3rd Order Butterworth bandpass between 1 and 40 kHz) was used to reduce effects of flow noise and click rate in the raw signal. Root-mean square pressure level SPL re 1µPa was measured for the filtered signal as:

$$SPL_{rms} = 10\log_{10}\left(\frac{1}{T}\int p^2\,(t)dt\right) \qquad \textit{Equation 5.2.}$$

Where time window T was defined either as a fixed window of 3.5 ms from click start time (SPL_{rms}), or a shorter, variable time window that contained 90% of the click energy ($\int p^2$) in the fixed time window ($SPL_{90\%}$). Both types of window were used for diagnostic purposes (Fig. 5B.5). All hydrophones were set the same sensitivity of -188 dB re 1V/µPa.

Because the duration of the 90% time window could slightly vary across clicks, $SPL_{90\%}$ was measured in conjunction with sound exposure level $SEL_{90\%}$. SEL is the energy flux density, and accounts for the duration of the signal. For the same signal duration $SEL_{90\%}$ is $SPL_{90\%}$ plus log-transformed duration:

$$SEL_{90\%rms} = 10\log_{10}\left(\int_0^T p^2(t)\,dt\right)$$

$$= 10\log_{10}\left(\frac{1}{T}\int p^2\,(t)dt\right) + 10log_{10}(T) \qquad \textit{Equation 5.3}$$

Peak-to-peak sound pressure level (SPL_{pp}) was also measured, as recommended by Madsen (2005).

5.3.3 Data structures and quality control

Three data structures were defined for analysis: 1) click level data ("click data"), 2) binned time series for 0.5 second bins ("binned data"), and 3) summary statistics for each audited buzz ("buzz data") (Table 5.1). Click data were binned in order to obtain a time-balanced sample for a fine-scale analysis of click rate while buzz data were used to compare and classify buzzes as proxies for prey capture attempts.

QUALITY FILTERING

Buzzes were filtered in two consecutive steps, "coarse filtering" and "fine filtering". Coarse-filtering aimed to maximise sample size of buzzes with a high overall detection rate while removing variability in the definition of what constituted a buzz. Coarse-filtered buzzes excluded: 1) buzzes with maximum binned click rate < 5 Hz, 2) buzzes with average binned click rate < 10 Hz, 3) very short duration (< 2 s) buzzes, and 4) buzzes with minimum depth < 20 m (to remove influence of near surface acoustics).

Fine-filtering was used in any acoustic analyses of AOL, and aimed to include only the highest quality detections, both of which were subject to variable SNR conditions. This more stringent filtering was also necessary to remove occasional false positives from other non-tagged whales that were included in the automated click detection. Data were first removed within each buzz on a bin-by-bin and click-by-click basis. The buzz was accepted if no more than 5% of buzz time and less than 50 clicks were removed. Clicks and time bins were filtered based on thresholds for the difference between raw and smoothed (Tukey's running median) values for SEL_{90} within each buzz. SEL_{90} was smoothed across both the click data and binned data (maximum click SEL_{90}). Values that were less than the binned or click-by-click median SEL_{90} were not included in the smooth. The resulting missing values were interpolated linearly as a function of time from the smoothed values. Clicks and bins were then removed if the raw SEL_{90} values were above +12 dB or below -6 dB of the smoothed value. Removals were asymmetric because false detections from non-tagged whales or reverberations tended to have a lower measured SEL_{90}. As a second criterion to level, time bins with average ICI < 15 ms were also removed.

Measurements and summary statistics are listed in Table 5.1. All data were time stamped in real time, relative to tag-on time, and audit end time. Time to end of buzz was used because the pause at the end of a buzz was typically a more clear-cut and easier to mark than start time on the click series. Audit start time and end time defined buzz duration.

DE-TRENDING APPARENT OUTPUT LEVELS TO ACCOUNT FOR CHANGES IN SOURCE-TAG ASPECT

SEL_{90} values were detrended within each tag record (SEL_{90DT}) to account for any differences in tag attachment location and hydrophone sensitivity both within and across each tag. SEL_{90DT} was used as the primary sound metric for analysis to measure relative energy in each click and for comparison to cumulative levels, while the non-detrended SPL_{pp} are shown for diagnostic purposes. Generalised additive models (GAM, library mgcv in r) were used to estimate across-tag and within-tag effects on SEL_{90} (~ Gaussian) by including tag id as a factor covariate and time since tag-on time as a non-linear (smooth) interaction term with tag id (to account for any changes in tag position). Depth was also included as a smooth covariate to account for differences in dive depth used by the whales between and within tag recordings. The complexity of the smooth for depth was limited to 5 knots, i.e. a hyperbole (Section 2.2.5). The model was fitted to all coarse-filtered binned data, and then used to predict SEL_{90} with 20 metres as a reference depth for each time since tag-on time in both click data and binned data. To obtain a detrended value (SEL_{90DT}), the raw SEL_{90} (for each click, or maximum in a bin) was subtracted the predicted value. Only fine-filtered data were used in model fitting and prediction.

5.3.4 Statistical analysis

Three analysis steps were designed to decompose any clustering of movement and echolocation tactics from direct effects of depth (pressure, light) across different prey encounters. First, buzzes were classified using an unsupervised clustering algorithm on summary movement variables that were *a priori* expected to be related to prey mobility and manoeuvrability rather than to the physical effects of depth (e.g. pressure or light conditions). Second, if the existence of multiple clusters was supported, their

vertical distribution was tested by modelling depth in a generalised estimating equation (GEE) with the movement classification as an explanatory variable. Third, the importance of the movement classification ('hunting strategies') as an explanatory variable for acoustic metrics ('echolocation tactics': maximum click rate, and AOL at maximum rate) was tested in regression models. To account for any non-linear effects of pressure on echolocation, depth was included as a non-linear smooth covariate in a generalised additive mixed model (GAMM). These models could capture any effects of pressure in conjunction with clustering of movement behaviour on echolocation across buzzes within the buzz data (hereafter termed "across-buzz models"). Finally, to measure within-buzz variation in echolocation rates, binned click rate was modelled using binned AOL, turning angle (pitch and heading) and depth as explanatory variables.

CLUSTERING OF BUZZES BASED UPON MOVEMENT METRICS

Gaussian mixture models (GMM) were used for model-based, unsupervised clustering of the input data (function Mclust, r-package mclust). Models were fit using the EM (Expectation maximisation) algorithm that optimizes the likelihood iteratively to estimate the hidden membership of each data point in the univariate distributions that make up the multivariate normal distribution. Model parameterizations and number of clusters were selected by maximising the Bayesian information criterion (BIC). Maximum number of tested clusters was selected by allowing a maximum of two clusters per individual (2x number of individuals). Input data were sub-sampled to achieve a balanced sample across individuals and thus avoiding any undue influence of individual data on clustering. Number of individuals included in the analysis was determined based upon having a sufficient number of coarse-filtered buzzes per tag (≥130). Input data were log-transformed as required to satisfy the normality assumption, and centred (mean removed) to improve GMM convergence (function 'scale' in r base package).

Movement metrics were selected to be proxies for prey density (Inter-buzz-interval IBI), overall prey mobility and subsequent energy expenditure (sum of 2-norm ODBA),

and prey mobility in three dimensions (vertical velocity, pitching variance, and heading variance) (Table 5.1). Sum rather than mean of ODBA was used in order to more fully describe the total locomotion cost during a buzz. Conversely, duration of the buzz was not included because it may have been more directly limited by the air volume available to the whale at depth (Wahlberg 2002). Similarly, rolling behaviour was not included to avoid any confounding effects of light on body posture (Appendix 5D).

DEPTH DISTRIBUTION OF MOVEMENT CLUSTERS

To test for any vertical stratification of movement clusters (proxy for movement strategies), buzz mean depth was modelled as a Gaussian response variable in a generalised estimating equation (GEE) with the estimated movement classification as a factor explanatory variable (function geeglm in r package geepack). Any non-linear depth effects were accounted for because the explanatory covariate of interest was a factor and therefore captured the mean differences in depth across clusters. Tag identifier was used as a panel variable to estimate average effects across individual, and first-order autoregressive (AR1) correlation structure was specified to estimate serial correlation in subsequent buzzes. Empirical ('sandwich') standard error estimates are reported which are robust to the working correlation assumption.

EFFECTS OF DEPTH AND MOVEMENT CLUSTERS ON ACOUSTIC METRICS

Acoustic summary variables for each buzz were selected as potential proxies for prey mobility (e.g. maximum click rate thought to proxy data rate, Table 5.1) and any relationship between target strength and AOL. Because ICI appeared to co-vary with AOL metrics (Appendix 5C), two different metrics were used to summarise apparent level in each buzz. First, detrended maximum click SEL_{90DT} at the time of the binned maximum rate was used to proxy apparent output at a presumed short range (Table 5.1). Second, the effect of click rate on SEL_{90DT} was removed by fitting a linear model within each buzz with SEL_{90DT} as a response variable and expected TL (see below: *within-buzz variation in echolocation rates*, Table 5.1) as a univariate explanatory variable. The estimated intercepts could then be used as rate-corrected SEL_{90DT} for each buzz. For diagnostic purposes, these models were fit to both SPL_{pp} and SEL_{90DT} as

Gaussian response variables. The model intercepts, coefficients and R-squared values (% of variation explained) were stored for each buzz for exploratory analysis.

GAMM:s were fit with maximum click rate, SEL_{90DT} at the time of maximum rate, and SEL_{90DT} intercept as Gaussian response variables. Depth was fitted as a non-linear explanatory covariate (max number of knots 5, Section 2.2.5) and movement classification and tag identifier as factor covariates. A first-order autoregressive correlation structure (AR1) was fitted to account for any serial correlation in the time series within each individual (function gamm in mgcv library in r). Type III Wald tests were carried out to estimate the importance of cluster in explaining variation in the acoustic metrics, after individual variability and depth had been accounted for.

WITHIN-BUZZ VARIATION IN CLICK RATE

Click-level and binned data were used to investigate variation in ICI and click rate as a function of apparent click level, movement and depth. The analysis addressed three different but not necessarily mutually exclusive hypotheses: 1) ICI:s track target range that is a correlate of output click level due to AGC, 2) higher sampling rates are required to increase temporal resolution during increased manoeuvring and turning rates, and 3) click production rates are limited by hydrostatic pressure.

The first hypothesis was explored on the basis that apparent output could scale linearly with on-axis source level of the whale. If the ICI of buzz clicks was being adjusted to target range and click output adjusted to concomitant TL, then spherical transmission loss (Eq. 5.1) would predict a logarithmic decrease in click output level. The existence of such a relationship was investigated within each buzz by comparing SPL_{pp} and SEL_{90DT} with expected TL that would occur if ICI equalled the two-way travel time (TWTT) of sound between the emission and reception of echo (speed of sound assumed a constant 1490 m/s). TL was calculated as

$$TL = 40 \log_{10}(\frac{ICI}{2} * 1490 \frac{m}{s}) \qquad \textit{Equation 5.2}$$

and was contrasted with apparent output per click (Appendix 5C).

Binned click rate was modelled as a Gaussian response variable in a GEE for a multivariate test of the different sources of variation in clicking rates within buzzes. Due to the large sample size, function geem in r package geeM was used for efficient computation of the likelihood (McDaniel et al 2013). Binned rather than click-by-click data were used to avoid confounding effects of ICI and sample size in the model. In contrast to across-buzz models, buzz identifier was used as a panel variable for the within-buzz models. First-order autoregressive (AR1) correlation structure was specified to account for serial correlation within buzzes. In order to avoid influence of click rate deceleration due to sound production constraints, only “on approach” bins were included from fine-filtered data (n=13374, 111 min). Approach was defined as increase in ICI within each buzz at 0.5 second temporal resolution. The change in ICI was measured by smoothing (fine-filtered) binned click rate within each buzz (Tukey's running median), and calculating difference in the smoothed ICI of subsequent bins.

Candidate covariates were selected to test the three different hypotheses. First, maximum SEL_{90DT} was included to test the hypothesis that ICI is adjusted to target range. This hypothesis would be supported if the model estimated a negative log-linear relationship between rate and click level. A log-link was included in the model to account for the expected non-linearity. Second, angular change in pitch, heading and roll were included to test whether click rate was adjusted to increase temporal resolution with faster changes in the target position and ensonified water volume. This temporal resolution hypothesis would be supported if there was a positive relationship between rate and turning angle for heading and pitch, but not with change in rolling angle. Finally, click rate could also be limited due to pressure effects on pneumatic sound production. Pressure scales linearly with depth and air volume decreases exponentially (Boyle’s law). Therefore, both linear and inverse functions of depth were compared in the full model, and the lowest AIC model was retained for inference.

5.4 Results

5.4.1 Data

A total of 1700 buzzes were audited, of which 1358 (from 334 min recording time) were included in the analyses ('coarse' data-filtering, Table 5.2). 298 buzzes recorded during tagging periods, incidental sonar exposures and experimental exposure periods were excluded. From baseline periods, a further 32 shallow buzzes (min depth < 20 m) and 9 buzzes were removed due to poor overall detection (<10 Hz average detection rate). Further 'fine' data-filtered buzzes (n=910, 210 min) excluded almost all of the buzzes from whale sw08_152a, an unusually noisy tag where flow noise dominated energy from clicks at high repetition rates (Fig. 5B.2 a), and from sw09_141a that was associated with other whales for most of the baseline data period (Fig. 5B.2 b) (Table 5.2).

All acoustic metrics of apparent click output were strongly linearly correlated with each other (Fig. 5B.1), and negatively correlated with the duration (ms) that contained 90% of click energy (Fig. 5B.5). R-squared values were 0.03 (SPL_{rms}~duration), 0.2 (SPL_{90}~duration), 0.15 (SPL_{pp}~duration), 0.03 (SEL_{90}~duration) and 0.18 (SEL_{90DT}~duration) when fitting the respective variables as Gaussian linear models to all fine-filtered click data. 90% energy window duration appeared to increase as a function ICI up to ~ 30 ms (Fig. 5B.6).

Hereafter, energy flux density SEL_{90DT} is used as the primary AOL metric in order to account for any confounding effects of click duration and to compare apparent output at varying time windows. Detrended values are shown over raw values in order to control for within- and across tag variation in AOL. Raw peak-to-peak values are given for diagnostic purposes and for cross-literature comparison following Madsen et al (2005). Median and 95% quantiles are reported for both sound metrics unless stated otherwise.

5.4.2 Clustering of buzzes based upon movement

Movement clustering included five movement summary statistics: IBI, sum of 2-norm ODBA, pitching variance and heading variance were log-transformed, while average vertical velocity was not transformed (Fig. 5.1). Six tag records that contained at least 130 coarse-filtered buzzes were sub-sampled randomly to obtain 130 buzzes each for cluster analysis (780 buzzes in total).

BIC selection retained a 4-cluster GMM with diagonal, equal volume and varying shape for covariance (Fig. 5.2). While BIC appeared to level off when the number of clusters exceeded four, the highest BIC value was attained at eight clusters. However, this result was dependent on the exact mixture model used and therefore deemed a likely spurious result.

The two movement clusters with the deepest median depths (1388 and 1029 m) had the lowest median total ODBA values (4.5 and 4.3, respectively). Within these two clusters, the other included more descending vertical speeds (median 1.7 [1, 2.4]). The longest median IBI was obtained for the shallowest cluster (265 m), and the shortest IBI for the second shallowest (404 m) which also had the highest median ODBA (27.4). Based on these results, the four clusters are hereafter termed as “shallow-sparse”, “mid-active”, “deep-descent” and “deep”, respectively (Table 5.3).

5.4.3 Depth distribution of movement clusters

Cluster was not a significant predictor of depth across individuals in a GEE with AR1 correlation structure (χ^2=4.51, p=0.21), though it was highly significant without a correlation structure (χ^2=90.8, p<0.001). Nevertheless, in the GEE with AR1 working correlation, shallow-sparse and deep clusters were estimated coefficients that were significantly different from each other based upon their empirical standard errors (Wald test statistic 16.0, p=0.038). These results indicate high variability in movement classification relative to high serial correlation in the foraging depth selection of the six tags included in the classification.

5.4.4 Effects of depth and movement cluster on AOL metrics

Depth was estimated with linear relationships in all three GAMM models (maximum click rate (Hz), SEL_{90DT} at the time of maximum rate, and SEL_{90DT} intercept), with the smooth of depth estimated one degree of freedom in each model (Fig. 5.3 a). Linear GEE models could therefore be fitted instead to estimate individual-average parameters, and similarly to GAMM:s, were specified 1st order autoregressive working correlation.

There was strong support for the inclusion of both depth and cluster ($\chi^2 > 54$, $p<0.001$) in the GEE model for maximum click rate across buzzes (Table 5.4). There was similarly strong support for depth for both SEL metrics ($p<0.001$; Table 5.4), but weaker support for cluster in the model for SEL_{90DT} at maximum rate ($\chi^2 =10.4$, $p=0.015$) and no support for cluster in the model for SEL_{90} intercept ($\chi^2 =5.8$, $p=0.12$) (Table 5.4). Both GAMM and GEE models appeared to fit well to the data and the Gaussian distributional assumption appeared reasonable (Fig. 5.3 a-b, Fig. 5E.2).

Maximum click rate was estimated to decrease by 1.2 Hz (SE 0.14) for every 100 m increase in depth. SEL_{90DT} at the time of maximum rate and and SEL_{90DT} intercept were estimated to decrease by respective -0.27 and -1.10 dB re 1 µPa for every 100 m increase in depth (Table 5.4). Removing the effect of depth, the highest maximum rate was estimated for the shallow-sparse cluster (57.3 SE=1.35) that was significantly different from the deep-descent and deep clusters (at 1% level) but not from the mid-active cluster ($p=0.37$, Table 5.4). Conversely, the lowest maximum rate was estimated for the deep-descent cluster that was estimated to be -8.61 lower on average across individuals than in the shallow-sparse cluster (SE 1.29).

5.4.5 Within-buzz variation in click rate

The relationship between expected TL (Eq. 5.2) and AOL (both SPL_{pp} and SEL_{90DT}) was highly variable between buzzes. TL and click level appeared to correlate strongly and positively in some buzzes (Fig. 5C.1 a), co-vary but with apparent deviations in others (Fig. 5C.1 b), correlate negatively in few cases (i.e. positive correlation with click rate, Fig. 5C.1 c) and not correlate at all in many other buzzes (Fig. 5C.1 d). However, the

relationship was estimated mostly positive within the fine data-filtered buzzes, and SPL_{pp} and SEL_{90DT} were estimated with similar coefficients (Pearson correlation coefficient 0.91, n=910) at medians and 95% quantiles of 0.28 [-0.09, 0.58] and 0.24 [-0.13, 0.51]. Expected TL explained a median 39.8% [0, 83.6] of variation in SPL_{pp} and a median 35.6% [0, 83.9] of variation in SEL_{90DT}. There was a negative trend in click level intercepts as a function of buzz mean depth (Pearson correlation -0.74 and -0.76 for SPL_{pp} and SEL_{90DT}, respectively at R>0.5, n=313) (Fig. 5E.1). A negative trend was also apparent in total number of detections, average click rate, max click rate and SEL_{90DT} (at max click rate) as a function of depth within coarse-filtered buzzes (n=1358) (Fig. 5.4). Decreased click rates and SEL_{90DT} in deep (> 750 m) compared to shallower buzzes were consistent within each deep-diving tag record (Fig. 5.4).

The GLM model for binned click rate with mean depth (pitchDiff + headDiff + rollDiff + $maxSEL_{90DT}$ + meanP) obtained 1740.8 units lower AIC than with inverse depth (mean buzz depth)$^{-1}$. Therefore, mean depth was retained for further GEE analysis. Based on robust variance estimates and Wald tests, all GEE estimates except rollDiff (p=0.374) were highly significant (p<0.004) (Table 5.5). Across-buzz average click rate was estimated to increase slightly with change in pitch and heading (by a factor of 1.024 and 1.005 for every 10 degree increase, respectively), and decrease with $maxSEL_{90DT}$ and depth (by a factor of 0.962 and 0.964 for every 1 dB and 100 m respective increase) (Table 5.5). However, the model appeared to over-estimate click rate at low observed values, and under-estimate rate at high observed values (Fig. 5.5). The under- and over-estimation was insensitive to the removal of outliers (identified using cook's distance), changing distribution (Gamma or Poisson) or link function (identity, inverse or log). Similarly, residual plots indicated that Normal distribution was appropriate for modelling rate (Fig. 5C.2). Estimates were somewhat improved by modelling rate within a GAMM (buzz id as a panel variable and ar1 correlation) that estimated the most complex smooths for $maxSEL_{90DT}$ and meanP (estimated degrees of freedom 3.928 and 3.923, respectively; Fig. 5C.3).

5.5 Discussion

This chapter aimed to resolve whether there were different movement patterns during prey encounters and whether terminal echolocation behaviour was related to these movement patterns or pneumatically limited sound production at depth. Decomposing the effects of movement clusters and depth (pressure) was essential to show whether sperm whales target different prey at different depths. There was clear evidence for a linear decrease in buzz click rates (both total and instantaneous) with depth that could not be explained by movement cluster alone (Figs. 5.3-4, Table 5.4), indicating a pressure effect on sound production. Clustering that included inter-buzz-interval (IBI) and movement summary statistics, but not acoustic metrics or depth, produced a depth-dependent classification of four buzz types (Fig. 5.2). "Shallow-sparse" and "mid-active" buzz types incurred the highest total movement costs (sum of 2-norm ODBA), while shorter IBI, lower ODBA and fewer changes in orientation implicated a denser distribution of less mobile prey at depth (>750 m) ("descent-deep" and "deep" clusters, Table 5.3). Although the cluster analysis did not explicitly account for autocorrelation, the estimated series of clusters was highly serially correlated, indicating that the classification identified depth-specific prey layers. After accounting for effects of pressure, click rates were significantly higher during shallow-sparse than other types of buzzes (Table 5.4), and also weakly correlated with pitching and heading variance (Table 5.5), indicating that higher acoustic sampling rates were used to track more mobile prey. These results corroborate results from Teloni et al (2008) that sperm whales engage in at least two different foraging strategies, but also illustrate potential echolocation limitations at depth and multiple movement tactics within a dive (multiple movement clusters, Fig 5.6).

5.5.1 Evidence for pressure-driven effects

Depth explained significant amount of variation in maximum click rate and both metrics for AOL across buzzes (Table 5.4, n=551 buzzes). All three relationships were estimated linear when accounting for movement cluster (GAMM models; Fig. 5.3 a). Similarly, AIC model selection did not support inverse transform of depth covariate for

models of maximum rate, which was expected if air density rather than volume was important for the adjustment of echolocation rate (Boyle's law). On the other hand, if a fixed air volume was required to produce each buzz click, an exponential decrease may be expected in the total number of clicks produced in each buzz as a function of depth. The data appeared to support this prediction (Fig. 5.4), although total number of clicks was not included in the formal analysis due to confounding effects of buzz duration and ODBA.

Sperm whales regular click trains are interspersed with pauses that are thought to relate to air recycling events (Norris and Harvey 1972). Similar to the reduction in number of buzz clicks here, Wahlberg (2002) showed that the time interval between subsequent pauses as well as the number of regular clicks decreased at depth (up to 1500 m). Although regular clicks were outside the scope of this chapter, it is conceivable that both usual and buzz click production are pneumatically limited and that some of the air volume available for a buzz is used in the preceding series of regular clicks.

Two metrics for AOL were used to correlate with depth and cluster: SEL_{90DT} at the time of maximum rate and SEL_{90DT} intercept. The former metric described AOL at a presumed short range to target, and the latter a level that was removed any effect of ICI. Both AOL metrics correlated negatively with depth (Table 5.4) despite parallel reduction in click rates (Fig. 5.4) that tended to be negatively correlated with apparent output (Fig. 5E.1). The trend appeared consistent within all deep-diving tag records included in the analysis (n=5) (Fig. 5.4). Madsen et al (2002) showed that on-axis output and frequency of sperm whale regular clicks were independent of depth up to 700 m. While their result may appear superficially contradictory, multiple factors could explain the discrepancy to the off-axis results here, such as the different depth range, or a narrowing of acoustic beam to retain optimal sonar capacity of on-axis clicks at depth. Sperm whale echolocation clicks are highly directional, and the backward directed beam can be dominated by the initial 'p0' pulse produced in the distal air sac (Zimmer et al 2005). Indeed, the sound pressure levels received by the tag may not only represent off-axis levels, but may also be filtered and beamformed by the body of

the whale (Johnson et al 2009). Similarly, Thode et al (2002) reported a change in the frequency content of regular clicks with depth but the source-receiver aspect was unknown and possibly variable. Therefore, the negative trend in the apparent click levels does not necessarily indicate that on-axis levels are compromised at depth, but could also relate to changes in sound propagation, such as cooling or compression of tissue, or changing of the sound beam as a function of depth.

5.5.2 Evidence for different prey types

Buzz clustering that included IBI and movement, but not depth or acoustic variables, produced a highly serially correlated and depth-dependent distribution (Fig. 5.6). Two clusters occurred predominantly during deep dives (>750 m): both “deep” and “deep descent” clusters consisted of short duration (median 6.5 and 7.4 s, respectively) and relatively low movement cost (sum of 2-norm ODBA) buzzes. The two clusters differed by vertical velocity, and deep descent buzzes were also observed in relatively shallow dives (<500 m, Fig. 5.6). The other two clusters “shallow sparse” and “mid-active” were on average more than twice as long in duration, but three to six times as high in movement cost than the deep buzz clusters. Of all clusters, mid-active buzzes were the longest in pursuit duration, produced at the shortest intervals and incurred the highest movement costs. These buzzes also included up to 11 peaks in binned click rate, indicating multiple re-approaches (Table 5.3).

These results corroborate previous research showing that that sperm whales engage in an active search-and-pursue strategy (Amano and Yoshioka 2003, Miller et al 2004, Aoki et al 2012), but also that the level of activity during prey encounters varied with depth. Increased movement effort indicated that more mobile prey were targeted at shallower depths (<750 m), such as muscular cephalopods or fish. Conversely at deeper depths (~750-1800 m) sperm whales appeared to either be selecting more prey items or foraging more densely distributed prey that did not require as much movement effort to catch. Furthermore, lower click rates suggest that high acoustic sampling rates were not necessary to detect prey at these depths.

This dichotomous hunting strategy appears consistent with stomach content analyses that show sperm whales can take both large quantities of fish (Martin and Clarke 1986), as well as likely more sedentary Gonatus fabricii in their spawning grounds (Santos et al 1999, Bjørke 2001, Simon et al 2003) in the North Atlantic. Mature female Gonatus fabricii loose ability to swim and float in the water column as part of ontogenic changes during breeding as their muscle tissue disintegrates and mantle and fins become gelatinous (Bjørke 2001). Given the high regional and individual variability in sperm whale diet (e.g. Evans and Hindell 2004), it is possible that a range of more sedentary cephalopods was taken at depth. These could include smaller cephalopods that are generally more bioluminescent, neutrally buoyant, slower swimming and less muscular than larger squids, and therefore likely easier to catch, as well as dead cephalopods that eventually sink to the sea floor (Clarke et al 1993, Clarke 1996, Whitehead 2003).

Based upon the expectation that a predator minimised energy expenditure for expected returns, the shallow prey types can be expected to contain more energy or other nutritional reward, such as protein contained in more muscular (and hence faster) prey species. Given that sperm whales must balance both their energy budget and diving metabolism, foraging at likely lower quality prey at deeper depths is likely to carry other advantages, such predictable and abundant prey patches (Teloni et al 2008). Sperm whales could also take advantage of aggregations of terminally spawning cephalopods, and prefer slower, more gelatinous, neutrally buoyant cephalopod species that are easier to capture, despite lower calorific value (Clarke et al 1993, Clarke 1996). A slow hunting strategy and selection of prey types that are easier to catch at depth may also be partly driven by breath hold diving where bursts of activity may incur costly anaerobic metabolites including carbon dioxide (Fahlman et al 2014).

It is possible that the vertical posture and speed during buzzing was influenced by the predator's vertical movement prior to the buzz, depending on whether the individual was capturing prey while transiting, or dedicated to search at a prey layer (descent, layer-restricted search and ascent; Chapters 3-4). These effects may have over-emphasised the descent-deep cluster as a separate movement strategy (Fig. 5.6).

It is also conceivable that whales switch hunting or echolocation strategy for the same species if the prey availability or detectability changed with depth. Sea floor might provide both physical shelter and refuge from acoustic detection. It would be important therefore to include both behavioural state and distance to sea floor in future analyses of prey encounter strategies.

5.5.3 Echolocation tactics

To better understand within-buzz dynamics of echolocation rates and subsequent maximum rate for each buzz, three sources of variation were considered in a regression model for binned rate: 1) AOL (potential log-scaling with target range), 2) temporal resolution and manoeuvring (proxy for prey mobility), and 3) pressure effects on pneumatic sound production. While there was large variability across buzzes, GEE model of time-binned click rate supported all three types of adjustment of click rate. As predicted, click rates increased with changes in pitch and heading, but not roll, and rates decreased both with increasing apparent level and depth (Table 5.5). Similarly, maximum click rate in a buzz, but not AOL metrics, were related to movement strategy with highest rates used during shallow-sparse and mid-active clusters (Table 5.3). These results indicate that higher acoustic sampling were used to track more mobile prey. In contrast, variation in AOL metrics was driven primarily by depth and click rate (Figs. 5.3-5.5).

SEL_{90DT} had the greatest magnitude of effect on click rates (Table 5.5), but the relationship between expected transmission loss (TL) and ICI was highly variable, flatter than expected and not always positive within buzzes (Appendix 5C). Such flat response could be expected if ICI was adjusted in terms of sensory volume rather than AGC in click output. Auditory evoked potential experiments indicate that echolocating odontocetes have the ability to discriminate and track clicks and echoes at high temporal resolution (e.g. 5-20 ms in Risso's dolphins Mooney et al 2006). Therefore, if sperm whales tracked individual echoes, their sensory volume per second would increase by a factor of 60 by decreasing ICI from 1 s to 15 ms alone. The reduction in sensory volume may have been partly modulated by click duration (90% energy

window length) that appeared to decrease with increasing AOL (Fig. 5B.5) and decreasing ICI (Fig. 5B.6).

If ICI corresponded to TWTT, the ranges to targets would have been 63-74 m at 0.1 s and 26-37 m at 0.05 s ICI (assuming 0-15 ms processing delay). These ranges would be relatively long compared to other studied beaked whales, delphinids and many bats that switch to terminal echolocation ~1 body length away from their target (Madsen and Surlykke 2013). Alternatively, sperm whales emit buzz clicks sparingly at the beginning of the buzz, and do not reach their capacity of click-by-click (echo-by-echo) discrimination until the temporal resolution is required. This hypothesis would be consistent with the apparent pneumatic limitations (Section 5.4.1).

Visual cues may also be important in detecting bioluminescent cephalopods or prey silhouettes against mid-water 'sky' (Fristrup and Harbison 2002). Clarke et al (1993) reported that almost 78% of the consumed cephalopod species in the stomachs of sperm whales off Azores contained luminous organs. If visual cues bore importance in finding prey, posture may be expected to change as a function of light conditions (depth). This prediction is descriptively supported by the movement data during prey encounters (Appendix 5D): sw05_199a, sw09_153a, sw10_147 and sw10_149a rolled on their side at ~90-120 degrees as click rates peaked during shallow dives.

Sperm whales are likely to use suction feeding mechanism similar to dwarf and pygmy whales (Bloodworth and Marshall 2005). Physiology of the uniquely shaped short, wide tongue of the sperm whale supports this hypothesis (Werth 2004). Not mutually exclusive, it has also been speculated that sperm whales regular clicks contain enough energy to debilitate prey acoustically (Norris and Møhl 1983, Møhl et al 2000), although recent experimental study could not induce stunning or disorientation in herring using a simulated click similar to a sperm whale regular click (peak frequency 18 kHz, received peak-to-peak level 193 dB re: 1 µPa, Benoit-Bird et al 2006). The active pursuit and multiple approaches to mid-water prey do not necessarily contradict the stunning hypothesis. Descriptively however, there was no clear movement or terminal echolocation signature that would have corroborated a

stunning hypothesis, such as a sudden decrease in movement effort following a more intense AOL (Appendix 5D). Furthermore, if AOL indeed scaled linearly with click output level, there was little support for a relationship between AOL and movement strategy per se.

5. References

Amano M, Yoshioka M (2003) Sperm whale diving behavior monitored using a suction-cup-attached TDR tag. Mar Ecol Prog Ser 258:291–295

Aoki K, Amano M, Mori K, Kourogi a, Kubodera T, Miyazaki N (2012) Active hunting by deep-diving sperm whales: 3D dive profiles and maneuvers during bursts of speed. Mar Ecol Prog Ser 444:289–301

Au W (1993) The sonar of dolphins. Springer-Verlag New York, US

Benoit-Bird KJ, Au WWL, Kastelein R (2006) Testing the odontocete acoustic prey debilitation hypothesis: No stunning results. J Acoust Soc Am 120:1118

Bjørke H (2001) Predators of the squid Gonatus fabricii(Lichtenstein) in the Norwegian Sea. Fish Res 52:113–120

Bloodworth B, Marshall CD (2005) Feeding kinematics of Kogia and Tursiops (Odontoceti: Cetacea): characterization of suction and ram feeding. J Exp Biol 208:3721–30

Boyd I (1997) The behavioural and physiological ecology of diving. Trends Ecol Evol 12:213-217

Clarke M (1996) Cephalopods as prey. III. Cetaceans. Philos Trans R Soc Lond B Biol Sci 351:1053–1065

Clarke MR, Martins HR, Pascoe P (1993) The diet of sperm whales (Physeter macrocephalus Linnaeus 1758) off the Azores. Philos Trans R Soc Lond B Biol Sci 339:67–82

Davis R, Jaquet N, Gendron D, Markaida U, Bazzino G, Gilly W (2007) Diving behavior of sperm whales in relation to behavior of a major prey species, the jumbo squid, in the Gulf of California, Mexico. Mar Ecol Prog Ser 333:291–302

Evans K, Hindell M (2004) The diet of sperm whales (Physeter macrocephalus) in southern Australian waters. ICES J Mar Sci 61:1313–1329

Fahlman A, Tyack PL, Miller PJO, Kvadsheim PH (2014) How man-made interference might cause gas bubble emboli in deep diving whales. Front Physiol 5:1–6

Finneran JJ, Mulsow J, Houser DS (2013) Auditory evoked potentials in a bottlenose dolphin during moderate-range echolocation tasks. J Acoust Soc Am 134:4532

Fristrup K, Harbison G (2002) How do sperm whales catch squids? Mar Mammal Sci 18:42–54

Gaskin D, Cawthorn M (1967) Diet and feeding habits of the sperm whale (Physeter catodon L.) in the Cook Strait region of New Zealand. New Zeal J Mar 1.2:37–41

Heithaus MR, Frid A, Wirsing AJ, Worm B (2008) Predicting ecological consequences of marine top predator declines. Trends Ecol Evol 23:202–10

Huggenberger S (2014) An acoustic valve within the nose of sperm whales Physeter macrocephalus. Mamm Rev 44:81–87

Johnson M, Aguilar de Soto N, Madsen PT (2009) Studying the behaviour and sensory ecology of marine mammals using acoustic recording tags: a review. Mar Ecol Prog Ser 395:55–73

Johnson M, Madsen PT, Zimmer WMX, Soto NA de, Tyack PL (2004) Beaked whales echolocate on prey. Proc R Soc B-Biological Sci 271:S383–S386

Johnson M, Madsen PT, Zimmer WMX, Soto NA de, Tyack PL (2006) Foraging Blainville's beaked whales (Mesoplodon densirostris) produce distinct click types matched to different phases of echolocation. J Exp Biol 209:5038–5050

Kawakami T (1980) A review of sperm whale food. Sci Reports Whales Res Inst 35:299-304

Madsen PT, Johnson M, Soto NA de, Zimmer WMX, Tyack P, Aguilar de Soto N (2005) Biosonar performance of foraging beaked whales (Mesoplodon densirostris). J Exp Biol 208:181–194

Madsen PT, Payne R, Kristiansen NU, Wahlberg M, Kerr I, Møhl B (2002) Sperm whale sound production studied with ultrasound time/depth-recording tags. J Exp Biol 205:1899–906

Madsen PT, Soto NA de, Arranz P, Johnson M (2013) Echolocation in Blainville's beaked whales (Mesoplodon densirostris). J Comp Physiol A Neuroethol Sens Neural Behav Physiol 199:451–69

Madsen PT, Surlykke a (2013) Functional convergence in bat and toothed whale biosonars. Physiology (Bethesda) 28:276–83

Martin A, Clarke M (1986) The diet of sperm whales (Physeter macrocephalus) captured between Iceland and Greenland. Mar Biol Assoc United Kingdom:779–790

Miller PJO, Johnson MP, Tyack PL (2004) Sperm whale behaviour indicates the use of echolocation click buzzes “creaks” in prey capture. Proc R Soc London Ser B Biol Sci 271:2239–47

Møhl B, Wahlberg M, Madsen PT, Miller L a, Surlykke a (2000) Sperm whale clicks: directionality and source level revisited. J Acoust Soc Am 107:638–48

Mooney TA, Nachtigall PE, Yuen MML (2006) Temporal resolution of the Risso’s dolphin, Grampus griseus, auditory system. J Comp Physiol A Neuroethol Sens Neural Behav Physiol 192:373–380

Naito Y, Costa DP, Adachi T, Robinson PW, Fowler M, Takahashi A (2013) Unravelling the mysteries of a mesopelagic diet: a large apex predator specializes on small prey (C Franklin, Ed.). Funct Ecol 27:710–717

Norris K, Harvey GW (1972) A theory for the function of the spermaceti organ of the sperm whale (Physeter catodon L). NASA Spec Publ 262:397

Norris K, Mohl B (1983) Can odontocetes debilitate prey with sound? Am Nat 122:85–104

Ridgway S, Carder D (2001) Assessing hearing and sound production in cetaceans not available for behavioral audiograms: Experiences with sperm, pygmy sperm, and gray whales. Aquat Mamm 27.3: 267-276

Santos M, Pierce G (2002) Additional notes on stomach contents of sperm whales Physeter macrocephalus stranded in the north-east Atlantic. J Mar Biol Assoc UK 82.03:501–507

Santos M, Pierce G, Boyle P (1999) Stomach contents of sperm whales Physeter macrocephalus stranded in the North Sea 1990-1996. Mar Ecol Prog Ser 183:281–294

Schnitzler H-U, Moss CF, Denzinger A (2003) From spatial orientation to food acquisition in echolocating bats. Trends Ecol Evol 18:386–394

Sergio F, Caro T, Brown D, Clucas B, Hunter J, Ketchum J, McHugh K, Hiraldo F (2008) Top Predators as Conservation Tools: Ecological Rationale, Assumptions, and Efficacy. Annu Rev Ecol Evol Syst 39:1–19

Simon MJ, Kristensen TK, Kinze CC, Tougaard S, Tendal OS (2003) Gonatus fabricii (Mollusca, Theuthida) as an important food source for sperm whales (Physeter macrocephalus) in the Northeast Atlantic. Sarsia North Atl Mar Sci 88:244–246

Slater GJ, Price S a, Santini F, Alfaro ME (2010) Diversity versus disparity and the radiation of modern cetaceans. Proc Biol Sci 277:3097–104

Teloni V, Mark JP, Patrick MJO, Peter MT, Johnson MP, Miller PJO, Madsen, Madsen PT (2008) Shallow food for deep divers: Dynamic foraging behavior of male sperm whales in a high latitude habitat. J Exp Mar Bio Ecol 354:119–131

Thode A, Mellinger DK, Stienessen S, Martinez A, Mullin K (2002) Depth-dependent acoustic features of diving sperm whales (Physeter macrocephalus) in the Gulf of Mexico. J Acoust Soc Am 112:308

Urick RJ (1984) Ambient noise in the sea. Peninsula Publishing, Los Altos, California, US

Wahlberg M (2002) The acoustic behaviour of diving sperm whales observed with a hydrophone array. J Exp Mar Bio Ecol 281:53–62

Werth AJ (2004) Functional Morphology of the Sperm Whale (Physeter macrocephalus) Tongue, with Reference to Suction Feeding. Aquat Mamm 30:405–418

Whitehead H (2003) Sperm whales: social evolution in the ocean.

Whitehead H, MacLeod CD, Rodhouse P (2003) Differences in niche breadth among some teuthivorous mesopelagic marine mammals. Mar Mammal Sci 19:400–406

Wisniewska DM, Johnson M, Beedholm K, Wahlberg M, Madsen PT (2012) Acoustic gaze adjustments during active target selection in echolocating porpoises. J Exp Biol 215:4358–4373

Ya Supin A, Nachtigall PE (2013) Gain control in the sonar of odontocetes. J Comp Physiol A Neuroethol Sens Neural Behav Physiol 199:471–8

Zimmer WM, Tyack PL, Johnson MP, Madsen PT (2005) Three-dimensional beam pattern of regular sperm whale clicks confirms bent-horn hypothesis. J Acoust Soc Am 117:1473–1485

5. Tables & Figures

Table 5.1 Measurements and summary statistics

Data set	Measurement	Description
Click data	SPL_{rms} SPL_{90}	Root-mean square pressure level SPL (dB rms re 1µPa) over a 3.5 detection window or 90% energy window (eq 5.3)
	SEL_{90} SEL_{90DT}	Sound exposure level / energy flux density (dB re 1µPa^2 s^{-1}) over 90% energy window (eq 2). DT indicates detrended value.
	SPL_{pp}	Peak-to-peak level (dB rms re 1µPa) over 3.5 ms window)
	duration	Duration of the 90% window for each click
	ICI (s)	Inter-click-interval; time since previous click detection (s)
Binned data	Click rate (Hz)	Click detection rate (Hz): number of detected clicks 0.5 s bin x2
	Max SPL_{pp} Max SEL_{90} Max SEL_{90DT}	Maximum click levels for respective sound metrics in 0.5 s bin SPL_{pp} in 0.5 s bin
	SEL_{cum} SEL_{cumDT}	Cumulative sound exposure level of click $SEL_{90\%}$ over 0.5 s bin Detrended (DT) cumulative sound exposure level of click $SEL_{90\%}$
	Pitch change Roll change Heading change	Shortest circular distance between the first and last samples of the bin (sampling rate 5 Hz)
	Depth	Mean depth (m) within the 0.5s bin
Buzz data	Start time, end time, duration (s)	Audited start time and end time from tag-on time in seconds, time interval between them defines buzz duration
	IBI (min)	Inter-buzz-interval: the time interval (h) between previous buzz (+any subsequent pause), and the start time of the next buzz
	Pause duration (s)	Time interval between the end of buzz and the start of next buzz or regular click train duration
	Max rate (Hz)	Maximum click rate, at 0.5 s bin resolution
	# max rate peaks	Number of peaks in buzz (from binned rate data)
	Max SEL_{90DT} at max rate	Detrended maximum SEL_{90} for the 0.5 bin that contained maximum click rate in the buzz
	SPL_{pp} and SEL_{90DT} and intercepts and coefficients	Intercept and coefficient estimates from the linear model with click level as a response variable and expected TL ($40\log_{10}(\frac{ICI}{2} * 1490\frac{m}{s})$) as an explanatory variable, fitted to the fine-filtered click data within each fine-filtered buzz
	ODBA	Sum of 2-norm ODBA over the buzz duration
	Vertical velocity	Signed depth difference between start and end of buzz divided by buzz duration
	Pitching and heading variance	Respective angular variances divided by buzz duration
	Depth	Mean depth (m) across 5 Hz depth values during the audited buzz

Table 5.2 Sample size and descriptive statistics by individual

	Sample size (# of buzzes)			IBI (min)	Dur (s)	Depth (m)			@ max click rate		
	audits	"coarse"	"fine"	median	median	median	min	max	Rate (Hz) median	SEL (dB) median	SEL (dB DT) median
sw05_196a	170	162	126	3.30	14.0	126.5	88.0	492.2	56	178.7	-8.4
sw05_199a	150	98	45	4.40	13.8	160.7	44.8	1309.2	48	175.5	-7.1
sw05_199b	201	172	134	1.06	7.1	1436.0	45.5	1859.0	34	171.3	-10.5
sw05_199c	245	131	107	0.75	5.9	1497.8	63.8	1753.7	36	168.1	-12.2
sw08_152a	47	26	1	3.83	11.8	110.9	70.1	701.2	50	159.4	-6.6
sw09_141a	58	11	4	8.94	12.6	162.6	22.6	589.3	44	161.8	-13.2
sw09_142a	55	46	22	3.48	16.4	348.3	59.2	631.6	52	175.2	-8.2
sw09_153a	65	64	58	3.86	16.7	130.4	59.1	349.7	54	171.3	-6.9
sw09_160a	261	236	183	1.39	8.6	1003.6	185.2	1625.9	42	173.5	-12.1
sw10_147a	155	142	56	0.71	18.0	462.6	80.1	669.9	56	170.3	-7.8
sw10_149a	92	80	34	4.06	11.7	236.1	46.7	1122.0	64	173.0	-4.5
sw10_150a	201	190	140	0.65	16.3	341.7	94.1	576.9	50	170.6	-7.7

Sample size: number of buzzes in each filtering category. Buzz summary statistics are defined in Table 5.1, and are calculated across all coarse-filtered buzzes (n=1351).

Table 5.3 Sample size and summary statistics for each cluster of buzzes based upon movement

Table 5.3 a) sub-sampled and coarse-filtered buzzes (n=780)

Cluster	196a	199b	199c	160a	147a	150a	Median [95% quantile] Depth (m)	Buzz dur (s)	Pause dur (s)	IBI (min)	ODBA (mean ∑)	Vertical v. (m/s)	pitch var. (rad $^{-1000}$)	head var. (rad $^{-1000}$)
Shallow sparse	86	20	9	28	42	24	265.3 [88.1, 1557.7]	13.6 [6.4, 24.7]	7.1 [0, 19.3]	3.7 [0.1, 60.9]	13.3 [6, 29.1]	-0.1 [-1.2, 0.8]	2.5 [0.3, 11.9]	3.4 [0.6, 25.8]
Mid active	33	1	4	10	60	89	404 [106.8, 691.3]	21.3 [9.9, 68.5]	3.3 [0, 771.9]	0.2 [0, 20.9]	27.4 [10.7, 81.1]	0.5 [-0.8, 2.5]	6 [0.9, 43]	11.1 [2.4, 33.9]
Deep descent	2	29	35	13	21	14	1029.2 [281, 1767.8]	7.4 [4.3, 31.6]	3.9 [0, 13.7]	0.6 [0, 43.2]	4.3 [2, 11.6]	1.7 [1, 2.4]	0.7 [0.1, 5.4]	2.3 [0.3, 39.6]
Deep	9	80	82	79	7	3	1388.3 [149.2, 1769.1]	6.5 [4.2, 13.0]	6.3 [0, 14.8]	0.7 [0.1, 4]	4.5 [1.8, 8.5]	-0.4 [-1.9, 1.2]	1.1 [0.1, 8.9]	1 [0.1, 10.9]

Table 5.3 b) sub-sampled and fine-filtered buzzes (n=551)

Cluster	196a	199b	199c	160a	147a	150a	Median [95% quantile] # max rate peaks	max rate (Hz)	SEL @max rate (dB DT)	SEL intercept (dB DT)	SEL coefficient (Hz / dB DT)
Shallow sparse	71	10	9	18	5	12	1 [1, 4]	56 [34.2,62]	-8.4 [-14.5, -1.4]	-18 [-37, -6.9]	0.3 [0, 0.5]
Mid active	22	1	3	6	30	72	2 [1, 11.35]	52 [37.3, 62]	-8.2 [-13.8, 3.0]	-15.1 [-36.7, -0.5]	0.2 [-0.1, 0.6]
Deep descent	2	26	31	12	14	12	1 [1, 6]	36 [24, 52.3]	-10.2 [-14.7, -6.1]	-25.2 [-42.7, -3.7]	0.3 [-0.1, 0.5]
Deep	7	62	63	59	2	2	1 [1, 2]	36 [29.7, 56]	-12.3 [-16.5, -5.8]	-29.6 [-43.1, -14.5]	0.3 [0.1, 0.5]

Table 5.4 GEE models for acoustic metrics of buzzes as a function of movement cluster and depth

	Coefficients				Type III tests	
Model for max rate (Hz)	Estimate	SE	Wald	p-value	χ^2	p
intercept	57.53	1.35	1803.72	0.000		
cluster					54.9	0.000
mid-active	-0.73	0.81	0.80	0.372		
deep-descent	-8.61	1.29	44.56	<0.001		
deep	-4.38	0.89	24.18	<0.001		
depth (x 0.01)	-1.18	0.14	72.31	<0.001	72.3	0.000
Model for SEL at max rate (dB DT)						
intercept	-7.52	0.43	308.4	<0.001		
cluster					10.4	0.015
mid-active	0.29	0.41	0.51	0.474		
deep-descent	0.26	0.37	0.48	0.487		
deep	-0.65	0.36	3.23	0.072		
Depth (x 0.01)	-0.27	0.03	62.41	<0.001	62.4	0.000
Model for SEL intercept (dB)						
intercept	-14.31	1.63	76.87	<0.001		
cluster					5.8	0.12
mid-active	2.78	1.44	3.72	0.054		
deep-descent	0.91	1.85	0.24	0.625		
deep	-0.39	0.93	0.18	0.675		
Depth (x 0.01)	-1.1	0.15	52.6	<0.001	52.6	0.000

GEE coefficient estimates and Wald tests are shown for each estimate, as well as Type 3 tests for each explanatory variable. Estimates are at response scale and shallow-sparse at zero depth is the intercept. Models were fit to sub-sampled and fine-filtered buzzes (n=551).

Table 5.5 GEE model for factors related to binned click rate

	Estimate	Model SE	Wald	p-value	Predicted (Hz)
(Intercept)	3.47	0.155	268	<0.001	32.26 [31.4, 33.1]
pitch change	2.41×10^{-2}	7.51×10^{-2}	3.24	0.001	0.78 [0.29, 1.29] / 10 deg
heading change	4.70×10^{-3}	14.4×10^{-3}	2.91	0.003	0.15 [0.05, 0.26] / 10 deg
roll change	-2.42×10^{-3}	25.9×10^{-3}	-0.89	0.374	-0.08 [-0.25, 0.10] / 10 deg
SEL_{90DT}	-3.83×10^{-2}	0.98×10^{-2}	-19.4	<0.001	-1.21 [-1.30, -1.11] / 1 dB DT
mean depth	-3.66×10^{-2}	2.10×10^{-2}	-17.1	<0.001	-1.16 [-1.26, -1.05] / 100 m

Model-based standard errors (SE) are shown, along with Wald-test static and p-values. Predicted values [+/- 2SE] are given for the intercept (all explanatory variables set to zero), and change from the intercept for the indicated unit increase in each explanatory variable.

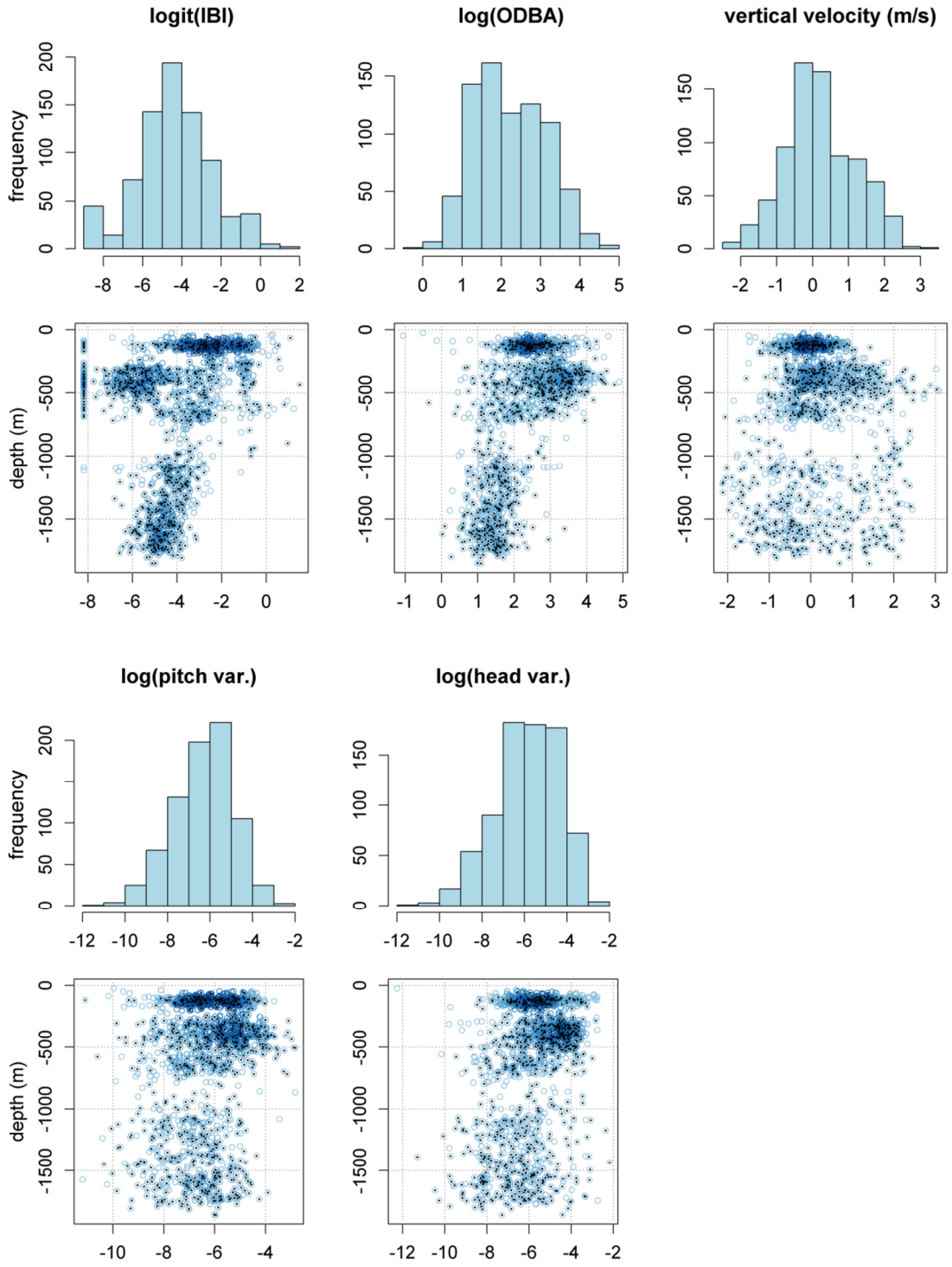


Figure 5.1. Summary statistics used for movement clustering

Data from coarse-filtered & sub-sampled buzzes (n=780, histograms and black circles) overlaid on all coarse-filtered data (n=1351, blue kernel density).

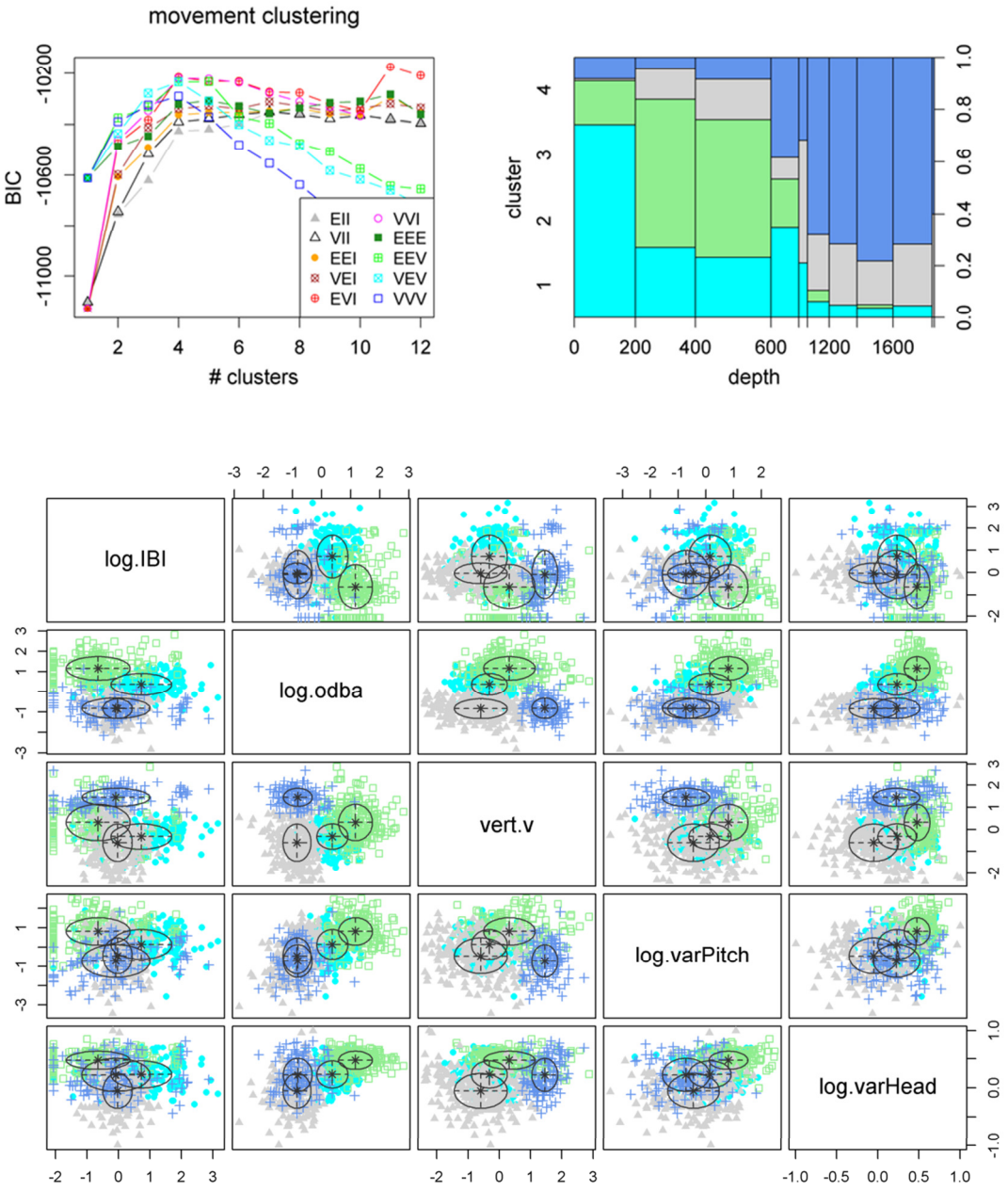


Figure 5.2 Clustering of movement metrics

Top Left: BIC selection of the number of clusters (x-axis) and parameterisations (colour-symbol combinations) of the GMM. Top right and bottom: clusters are coloured by median depth of the clusters (Table 5.3) from shallow to deep: cyan (1: shallow-sparse), green (2: mid-active), grey (3: 'deep-descent') and blue (4: 'deep'). Coarse-filtered and sub-sampled buzzes are shown (n=780, Table 5.3).

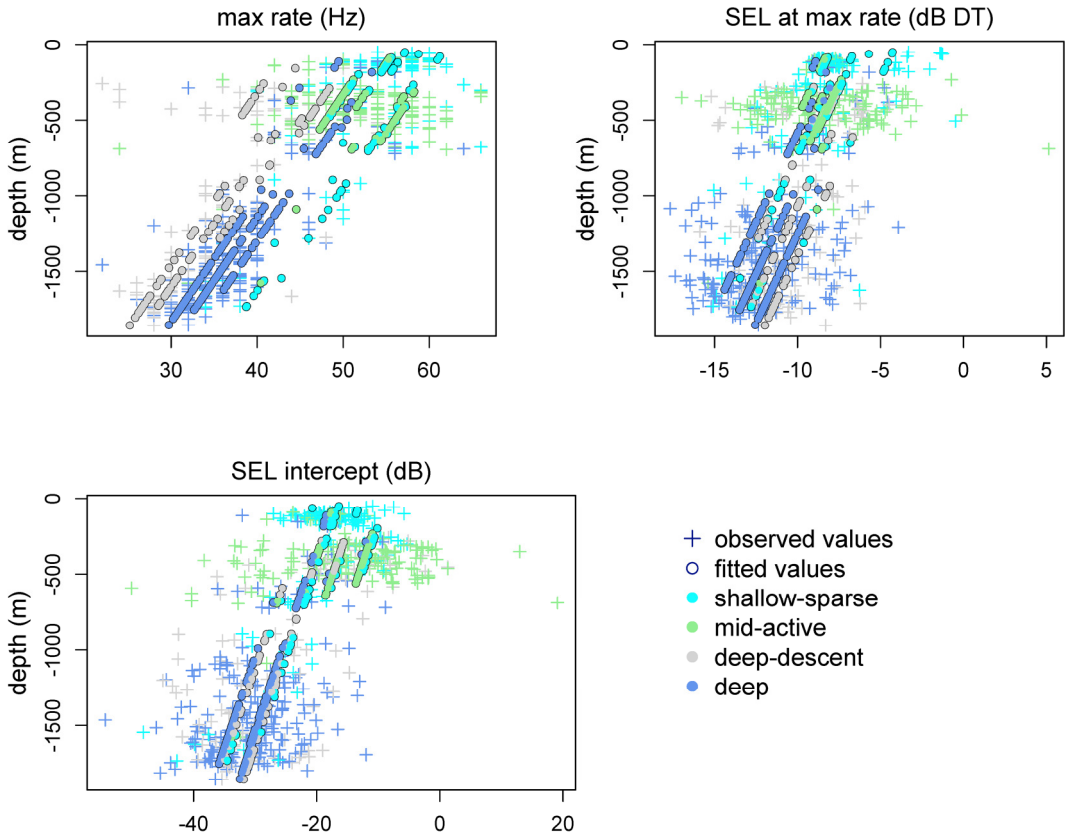


Figure 5.3 a) GAMM fitted and observed values for acoustic metrics as a function of movement cluster and depth

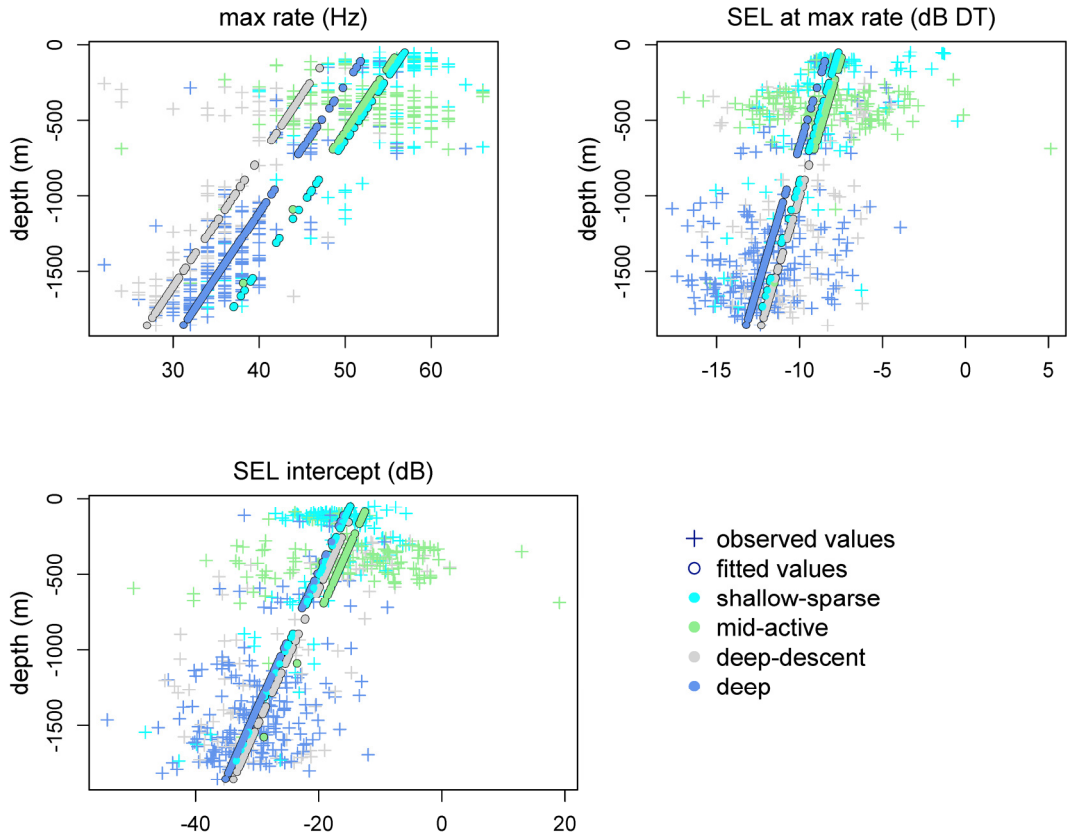


Figure 5.3 b) GEE fitted and observed values for acoustic metrics as a function of movement cluster and depth

Figure 5.3 Models for acoustic summary variables

Models were fit to fine-filtered and sub-sampled buzzes (n=551, Table 5.3). Crosses show observed values, and solid circles fitted values. Figure a) shows GAMM *acoustic metric ~ smooth(depth) + factor(cluster) + factor(tag id)* estimates that illustrate the linearity of the relationship and inter-individual variability when parameters are estimated separately for each individual. GAMM:s explained 80, 39.5 and 52.2% (adjusted R^2) of max rate, SEL_{90DT} at max rate, and SEL_{90} intercept respectively. Figure b) shows GEE *acoustic metric ~ depth + factor(cluster)* estimates that illustrate the individual-average differences between clusters. Respective GEE model estimates are shown in Table 5.4.

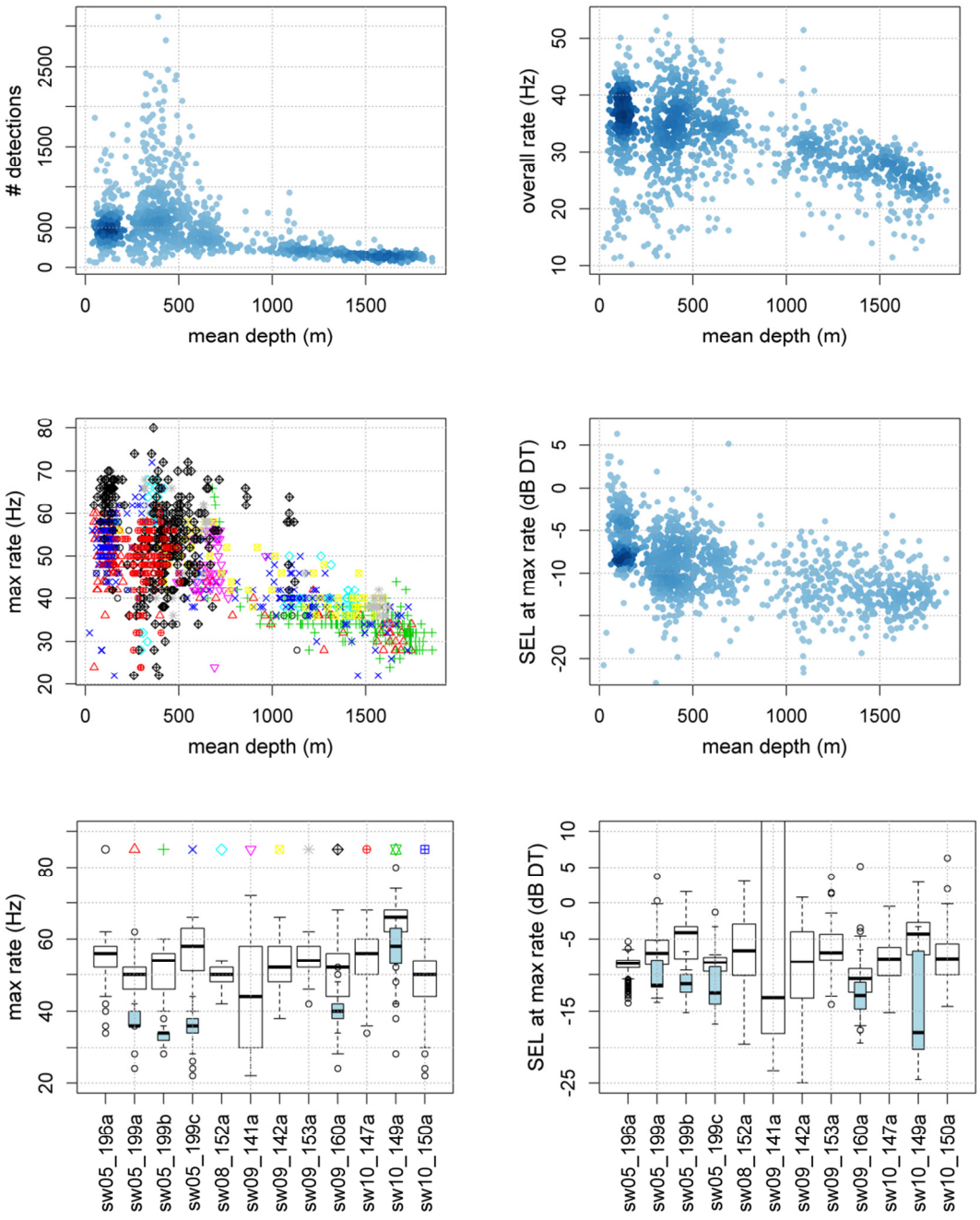


Figure 5.4. Buzz click rates as a function of depth

Number of detections, detection rate, maximum rate (at 0.5 s bins), and detrended maximum click SEL (dB) at the time of maximum rate. Summary statistics are shown for all coarse-filtered buzzes (n=1358), except 8 outliers for SEL at the bottom right figure. Bottom right: four SEL_{90DT} values that exceeded 10 dB in sw09_141a omitted. Blue colour scales show local kernel density of points (darker: denser). Box plots show deep (>750 m, blue) and shallow (<750 m) buzzes separately.

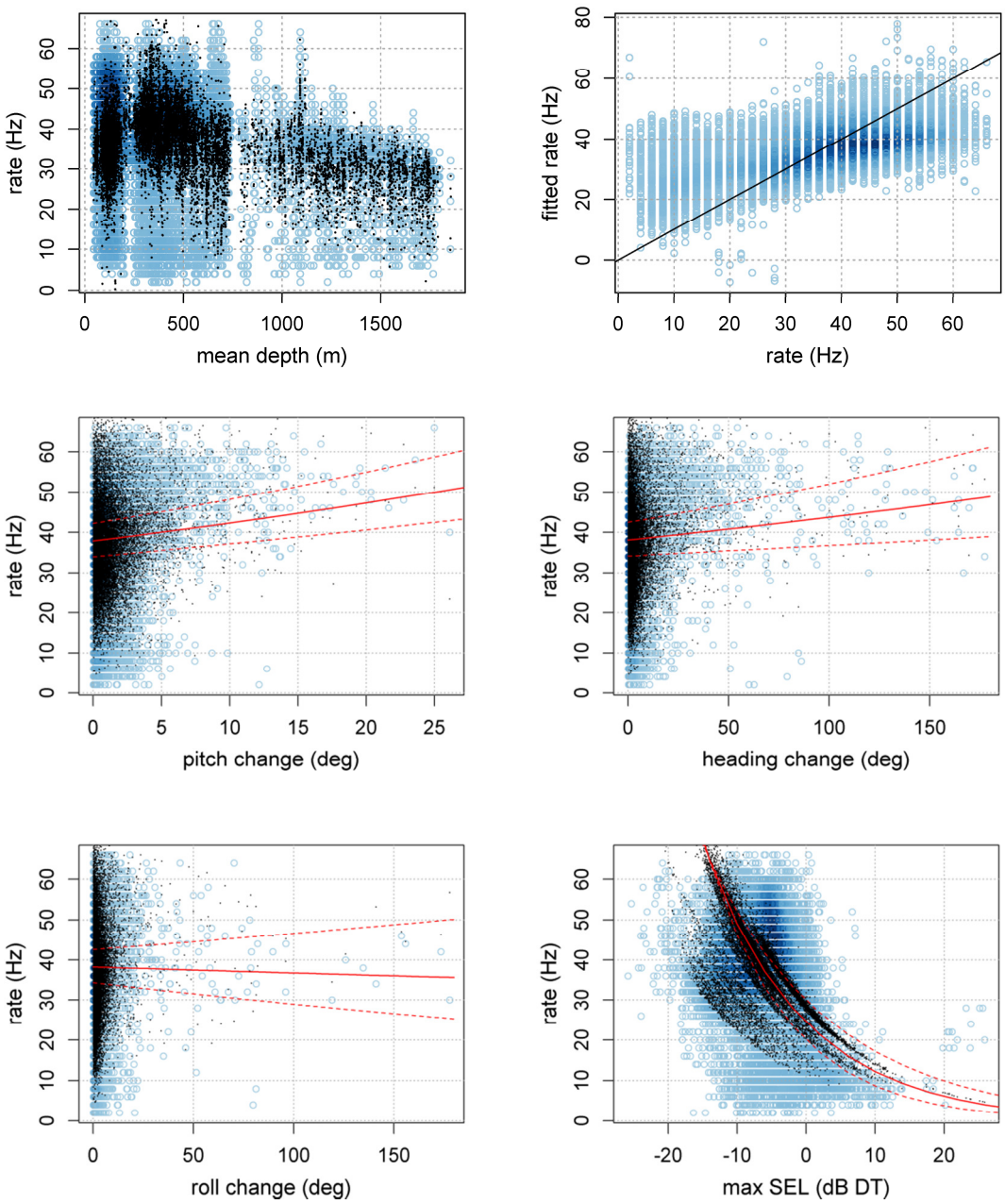
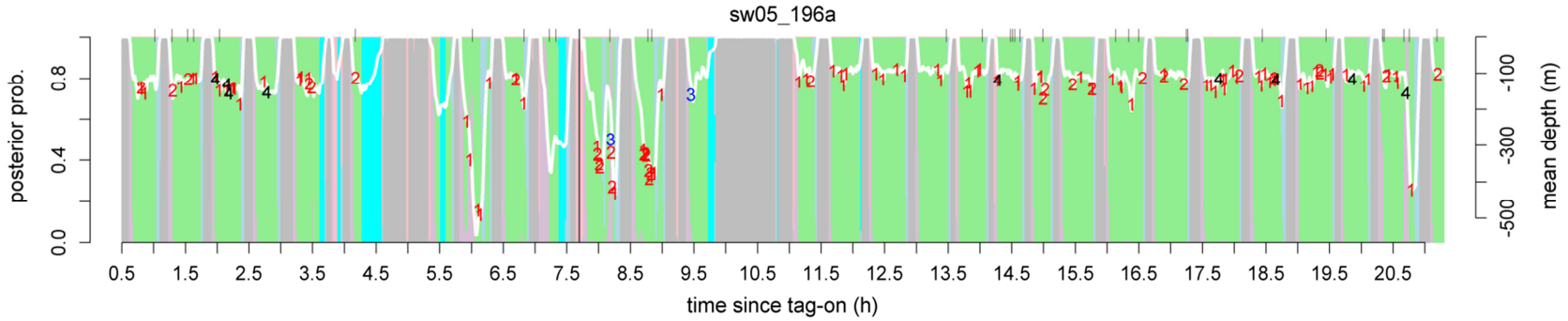
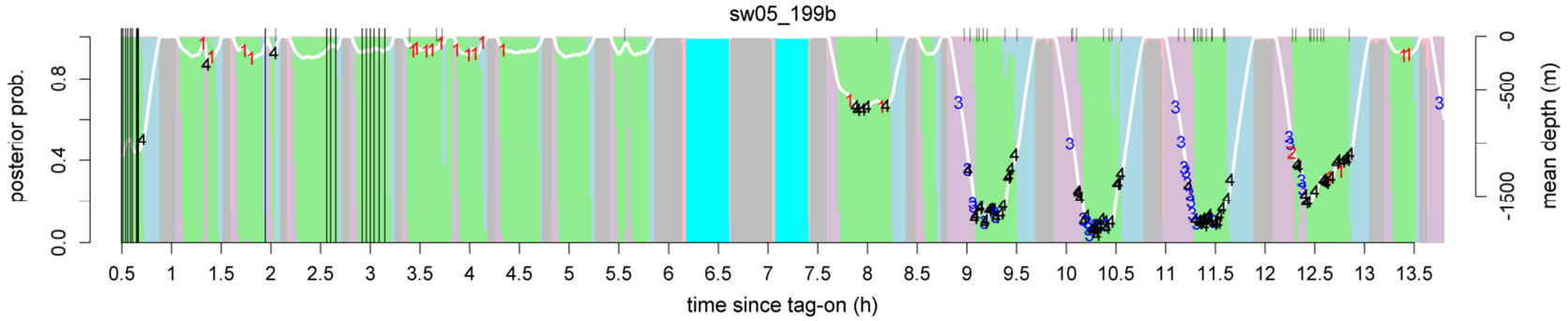
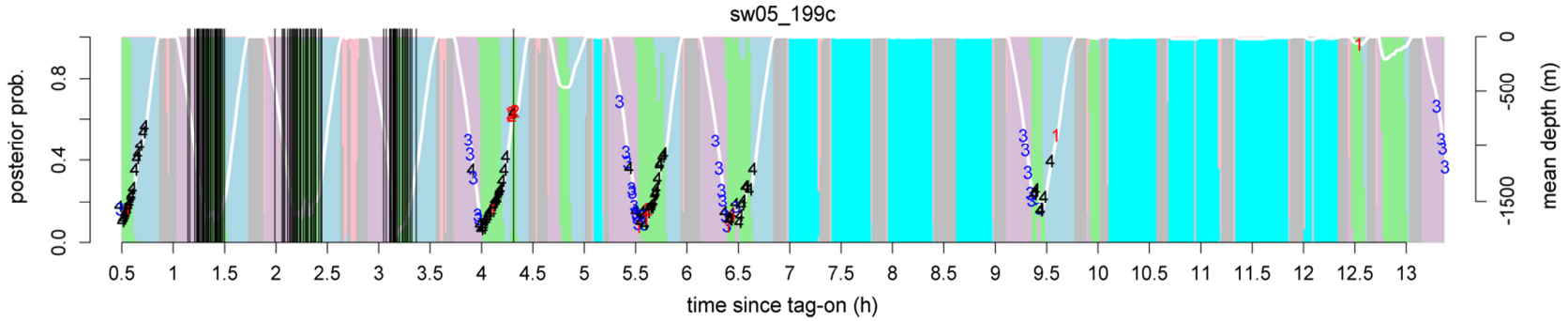


Figure 5.5 Observed and fitted values for binned click rate

Black: fitted values overlaid with blue: observed as a function of each explanatory variable in the Poisson GEE model for fine-filtered binned and "on approach" rate (n=111 min). Bottom right: fitted rates from the same model as a function of observed rate. Red: predicted values +/- 2x SE as a function of each explanatory variable, when all other explanatory variables were set to their median values.


sw05_196a
posterior prob.
0.0
0.4
0.8
mean depth (m)
-100
-300
-500
0.5 1.5 2.5 3.5 4.5 5.5 6.5 7.5 8.5 9.5 10.5 11.5 12.5 13.5 14.5 15.5 16.5 17.5 18.5 19.5 20.5
time since tag-on (h)


sw05_199b
posterior prob.
0.0
0.4
0.8
mean depth (m)
0
-500
-1500
0.5 1 1.5 2 2.5 3 3.5 4 4.5 5 5.5 6 6.5 7 7.5 8 8.5 9 9.5 10 10.5 11 11.5 12 12.5 13 13.5
time since tag-on (h)


sw05_199c
posterior prob.
0.0
0.4
0.8
mean depth (m)
0
-500
-1500
0.5 1 1.5 2 2.5 3 3.5 4 4.5 5 5.5 6 6.5 7 7.5 8 8.5 9 9.5 10 10.5 11 11.5 12 12.5 13
time since tag-on (h)

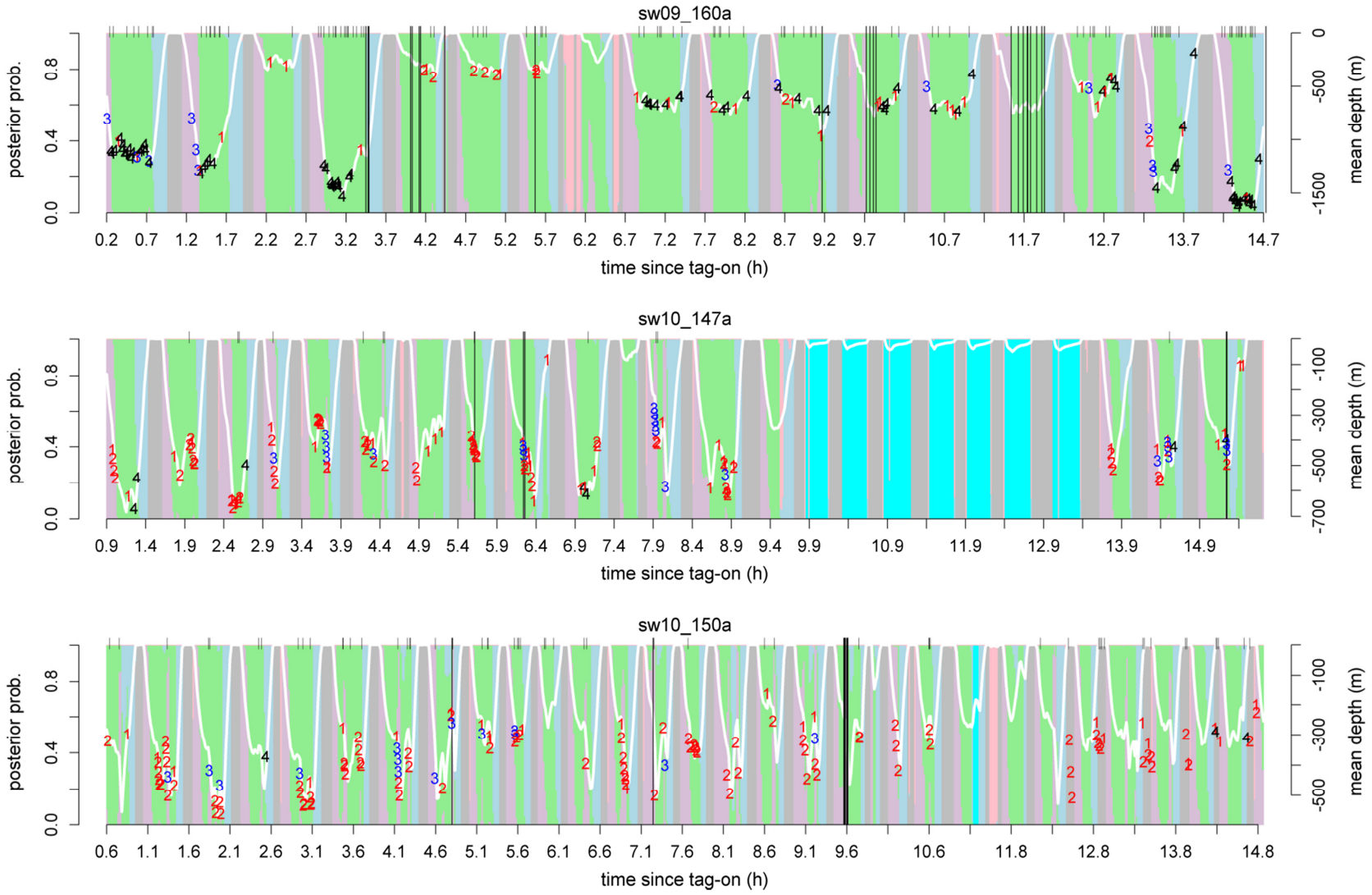

sw09_160a
posterior prob.
0.0
0.4
0.8
mean depth (m)
0
-500
-1500
0.2 0.7 1.2 1.7 2.2 2.7 3.2 3.7 4.2 4.7 5.2 5.7 6.2 6.7 7.2 7.7 8.2 8.7 9.2 9.7 10.7 11.7 12.7 13.7 14.7
time since tag-on (h)
sw10_147a
posterior prob.
0.0
0.4
0.8
mean depth (m)
-100
-300
-500
-700
0.9 1.4 1.9 2.4 2.9 3.4 3.9 4.4 4.9 5.4 5.9 6.4 6.9 7.4 7.9 8.4 8.9 9.4 9.9 10.9 11.9 12.9 13.9 14.9
time since tag-on (h)
sw10_150a
posterior prob.
0.0
0.4
0.8
mean depth (m)
-100
-300
-500
0.6 1.1 1.6 2.1 2.6 3.1 3.6 4.1 4.6 5.1 5.6 6.1 6.6 7.1 7.6 8.1 8.6 9.1 9.6 10.6 11.8 12.8 13.8 14.8
time since tag-on (h)

Figure 5.6 Movement clustering overlaid with dive profile and behavioural states estimated in Chapter 2.

Numbers indicate estimated clusters 1 (red): Shallow-sparse, 2 (red): mid-active, 3 (blue): deep-descent, 4 (black): deep (n=780). Vertical lines show timing of buzzes that were coarse-filtered, and top rug plot shows buzzes that were excluded due to sub-sampling. Background shading shows proportion of posterior probabilities for state (1-6; Chapter 2).

6 SYNTHESIS

6.1 Summary of work

The objective of this thesis was to quantify biologically relevant behavioural changes for a deep-diving marine mammal using sperm whales in their foraging ground as relatively simple model system.

Passive acoustic monitoring of echolocation clicks was used as the primary tool to measure changes in foraging behaviour, both remotely using towed hydrophones (Chapter 2) and onboard animal-attached digital acoustic recording tag (DTAG) (Chapters 3-5). Sperm whale 'regular clicks' and 'buzz clicks' were interpreted as indicators of searching for prey and prey capture attempts, respectively (Section 1.4.2). Time series models (generalised linear models GLM, generalised additive models GAM, and generalised estimating equations GEE) were used throughout for hypothesis testing while accounting for multiple sources of variation in response data. A biologically informed hidden state model was developed to estimate functional-behavioural time budgets of tagged sperm whales from biologging data (presence/absence of echolocation clicks, changes in pitch and depth) (Chapter 3). Although buzz rate and ODBA were not included in the state estimation, the time series of estimated states succeeded to capture natural variability in both buzz rate and ODBA across functional behavioural states (Chapters 3-4).

In Chapter 2 "effects of whale-watching", a range of acoustic metrics calculated for focally followed whales were tested for change before, during and after viewing by whale-watching boats and as a function of distance to the research vessel in the Kaikoura Canyon. While small changes in echolocation patterns were detected, these were within the behavioural range observed across the sampled individuals and spatiotemporal environment. Apparently large individual and contextual variability relative to sample size precluded detection of statistical significant changes in acoustic metrics most closely related to prey encounters. Distance to the research vessel correlated with buzz detections, but this was likely due to the recording system's ability to detect buzzes rather than a change in buzz production by the whale's themselves.

Chapter 3 “effects of tagging” and Chapter 4 “effects of sonar” utilized DTAG data from the controlled exposure experiment (CEE) study conducted by the Sonar, Sea Mammals, Safety (3S) collaboration in Norway to quantitatively contrast behavioural changes during tagging operations, experimental sonar, incidental sonar, and experimental negative and positive controls. Chapter 3 focused upon the development of a state-based analytical approach to detect changes in foraging behaviour, which was applied in both Chapters 3 and 4. The developed hidden state model was first used to estimate time series of biologically informed, functional states for all tagged whales (n=12). The functional-state time series was then used as input data with uncertainty to test hypotheses for behavioural change during the different exposures: 1) change in functional behaviour state, 2) change in proxies for foraging success (presence/absence of buzz), given a state, and 3) change in a proxy for locomotion cost (overall dynamic body acceleration ODBA), given a state.

In Chapter 3, hidden state models succeeded to estimate both very clearly defined functional states (layer-restricted search behaviour, drifting) and states that were less defined both in terms of a-priori expectation of a behavioural function and model specification (non-foraging silent active state). The random walk model for depth supported a depth-dependent vertical step length distribution, with both faster descent and ascent speeds at greater depths. Despite individual variability, there was enough statistical power to detect increased switching to surfacing and non-foraging silent active states in the presence of the tag boat than during post-tagging baseline periods. Time-decaying models of tagging effects were not retained in model selection, indicating a short-term effect that ceased immediately after the tag boat left the whale.

Chapter 4 investigated behavioural effects of exposure to sonar and playback of killer whale calls. This analysis benefited from the finding that tagging effects were limited to the beginning of the DTAG records (Chapter 3), and that costs could be quantified relative to the pre-exposure baseline period that followed the tagging phase. Similar to the response during tag boat presence, whales increased time allocation to a non-foraging ‘silent active’ state both during a stationary playback of

killer whale (*Orcinus orca)* sounds and experimental transmission of 1-2 kHz active naval sonar from an approaching vessel. Individuals switched to the silent active state near the onset of 1-2 kHz sonar transmissions at a wide range of received sound pressure levels (SPL 131-165 rms re 1µPa). A non-parametric analysis on time allocation revealed that the state-transitions also translated to changes in a dive-scale time budget.

Chapter 5 "linking buzzes to prey" combined both movement and acoustic DTAG data to more closely examine potential clustering in movement and sensory strategies during buzzing, and their vertical distribution. There was strong support for both vertical distribution of movement strategies possibly related to differences in prey at depth, but also likely pressure effects in pneumatic sound production evidenced by a clear linear reduction in buzz click rates as a function of depth. Nevertheless, not all variation in within-buzz click rate could be explained by depth alone; a weak correlation with pitching and heading variation, but not roll, suggested that higher echolocation clicking rates were used to pursue more mobile prey. Depth distribution of movement strategies indicated that sperm whales engage in an active pursuit of more mobile, and perhaps more nutritionally rewarding, prey at relatively shallow depths (<700 m).

These results illustrate that sperm whale foraging behaviour can be disrupted in response to a range of anthropogenic activities, and chronic exposure may be of particular concern for areas of high productivity such as the Kaikoura Canyon. The potential biological significance of the relatively short-term change in functional time budget to 1-2 kHz sonar was highlighted by the near-identical behavioural responses to playbacks of mammal-eating killer whale sounds. Nevertheless, sperm whales at the high-latitudes studied here are also shown to engage in a wide range of foraging strategies, suggesting that this generalist predator may be relatively plastic in terms of responses to anthropogenic disturbance.

6.2 Comparison of disturbance stimuli and behavioural effects

6.2.1 Severity and temporal scaling

Behavioural changes ranged from apparently benign changes in echolocation patterns in the presence of research and whale-watching vessels to likely more biologically significant costs in proxies for time and energy during 1-2 kHz active sonar and mammal-eating killer whale sound playbacks. Nevertheless, direct comparison of their severity in terms of potential fitness consequences is challenging. Here, I use an updated version of the original Southall severity scoring table from Miller et al (2012) (Appendix 6A) to summarise exposures and behavioural effects across Chapters 2-4. Southall (2007) defined the ordinal (0-9) severity scale to qualitatively review the links between documented behavioural responses to anthropogenic noise and population consequences within the PCAD framework (NRA, 2005). Their scale differentiated between behavioural changes that were likely to be relatively 'minor or brief' (scores 0-3) from behavioural effects that were deemed to impact foraging, reproduction or survival (scores 4-6) and vital rates (scores 7-9).

The table from Miller et al (2012) is organised into types of behavioural effects (avoidance, changes in locomotion etc.) in Table 6.1, and my personal judgment (non-consensus) scoring for each type of exposure are given in Table 6.2. The scores aim to follow the Miller et al (2012) consensus panel definitions for severity that was in part based upon the longevity of behavioural response relative to the duration of exposures. 'Brief and 'minor' behavioural responses were defined shorter than the exposure duration, while 'moderate' behavioural responses typically stopped as the exposure stopped. A response that continued beyond the end of exposure was considered 'prolonged' (Miller et al 2012). The exact severity score is therefore a somewhat subjective decision, but provides an explicit expression for any future comparison across studies.

Behavioural effects were studied in response to three types of vessels that followed focal whales both at surface and underwater: 6-m motored research vessel 'Titi' used for following and passive acoustic monitoring of a focal whale (Chapter 2), similarly

small (5-8 m), but rigid-hulled inflatable boat used to tag and re-approach whales for photo-identification (Chapter 3), and much larger (55 m) 'Sverdrup' that towed either silent or active sonar source (Chapter 4). In addition, effects of commercial whale-watching (ww) boats (17-18 m jet-engine catamarans) that typically restricted their attention to only surfacing whales were investigated in Chapter 2. Titi and commercial ww-boats followed a similar code of conduct near the whale at surface (New Zealand Marine Mammal Protection Regulations 1992), including approach from behind the whale, but Titi generally remained further away from the focal whale, especially when giving priority to the commercial fleet (ww-boats, helicopters and flights). All the four types of vessels could approach and re-approach the same whale multiple times, which may have resulted in variable degrees of sensitization or tolerance to their continued presence (Bejder et al 2009). The 3S experiments (Chapters 3-4) were conducted on a once-a-year basis on likely different individuals, while Titi and whale-watching boats encountered same individuals on a monthly and annual basis (Chapter 2). Consequently, comparable exposure durations are difficult to determine across the different vessel behaviours.

Tag boat was the only type of vessel that attempted to approach the focal whale closer than 50 m, and appeared to elicit the most severe response of all types of vessel operations that were studied (Table 6.1). No reduction in foraging effort could be detected in response to the focal follow vessel Titi, nor sonar source vessel Sverdrup. While these results might suggest that the closest approach distance is an important determinant of behavioural responses, direct comparison is also limited by the absence of control data for Titi (responses were assessed based upon the closest range to the whale at each surfacing, Section 2.3.4). In Kaikoura, land-based observations were carried out at the time of the whale-watching study (Markowitz et al 2011). Integration of these data with the acoustic and surface behaviour of the focal whales could improve the sample size and allow more conclusive estimation of effects for the research vessel.

The sperm whale had a common response signature to experimental LFAS, killer whale sound playbacks, and presence of tag boat, which consisted of whales spending

more time in a non-foraging silent active state, and less time in the surface state. Foraging is a key life function, and changes in foraging behaviour have been reported in a range of marine mammals in response to both natural predators (Wirsing et al 2008) and anthropogenic activity (Lusseau and Bejder 2007, Nowacek et al 2007, Tyack 2009). Modifications in echolocation patterns (Chapter 2) and in key behavioural states (Chapters 3-4) are qualitatively consistent with previously documented behavioural disturbances in sperm whales (e.g. whale-watching, Richter et al 2003, tagging and seismic airguns Miller et al 2009). Although a key life-function, changes in foraging behaviour are not always likely to impact the growth and body condition of the whales.

While response to the tag boat was estimated to cease immediately after the exposure stopped – and thus clearly fell within the 'moderate' severity category – responses to both LFAS and killer whale sound playbacks were estimated to last until 8 and 19 min into post-exposure (Table 4.5). Given the relatively shorter exposure duration of killer whale playback compared to LFAS, the total duration of both behavioural responses was estimated at ~ 40 min. While the difference in the estimated post-exposure effect seems relatively small (~10 min), this difference should be interpreted with caution as killer whale playbacks were conducted towards the end of most tag records. The duration of the DTAG deployment may have precluded the detection of longer post-exposure effects for killer whale playbacks. In contrast, sperm whales appeared consistently to resume to post-tagging baseline foraging behaviour relatively soon after LFAS exposures ended (Fig. 3D.3).

It therefore seems reasonable to conclude that while sonar and tagging procedures can trigger similar behavioural responses as simulated presence of a presumed natural predator, whales were able to reduce their perceived risk upon the vessels' departure (Section 4.4.3). Unlike mammal-eating killer whales that hunt silently (Deecke et al 2005), whales may be able to monitor vessels at surface due to engine noise. Indeed, this might explain why whales might extend the duration of first silence in the beginning of their dive in the presence of whale-watching vessels in Kaikoura (Chapter 2). With these caveats and apparent high variability in response intensity across

experiments (Fig. 3D.3), I concluded that the maximum severity for an observed response to both LFAS and killer whale playbacks was unlikely to impact vital rates, and therefore scored as 'moderate' (Table 6.2). These results highlight that temporal and spatial scaling is key to understanding the consequences of behavioural disturbance. The documented responses would indicate the potential for more biologically relevant consequences if animals were exposed to these stimuli frequently or over long durations.

It is worth emphasising that the analytical framework developed in Chapter 3 was especially designed for detecting "cessation of foraging", and was not designed to detect other potential types of behavioural changes (e.g. avoidance, change in locomotion or orientation). The lack of these effects in Table 6.1 is mainly due to the lack of testing for them. Furthermore, it is possible that had we tagged whales in Kaikoura, similarly severe responses could have been detected in foraging behaviour. Indeed, any silent dives of focally followed whales in Kaikoura would have likely resulted in poor tracking and incomplete dives, and thus would have been discarded from analysis. This is an important caveat of passive acoustic tracking, and is especially highlighted by the estimation of non-foraging silent active state as the "disturbed state" in Chapters 3 and 4.

6.2.2 Implications for conservation

Impacts on cetacean populations can be difficult to detect at a critical stage (Taylor et al 1993), and behavioural effects may provide the first indicators or mechanistic clues for such a numeric population-size response. From the 'ecology of fear' point of view, non-lethal but chronic effects of predators such as increased vigilance can impact population even more than lethal effects (Brown et al 1999). Sperm whale foraging behaviour can be altered in response to a range of short-term anthropogenic signals (Table 6.1), and their combined effect may be chronic reductions in foraging rates in areas where anthropogenic activities are frequent. Gulf of Mexico and the Mediterranean are examples of semi-enclosed seas that are both important areas for sperm whale and energy exploration, fishing, shipping, and naval testing (Miller et al

2009). Deep water canyons that are situated in the coastal zone, such as the Kaikoura canyon, are particularly important habitats due to their high productivity (Section 2.4.3), but may also pose increased anthropogenic disturbance (e.g. whale-watching) and other risks to whales (e.g. vessel collisions, chemical and noise pollution).

The detected behavioural changes in Chapters 2-4 were acute and short-term, but the key conservation question is whether these behavioural changes could persist over longer or repeated exposures. The severity of observed behavioural responses to tag boat presence was judged to be 'moderate' while the response to viewing by whale-watching vessels 'brief' (Table 6.1). However, the potential for a chronic exposure of sperm whales to tagging efforts is likely small, while sperm whales in Kaikoura Canyon are exposed to whale-watching year-round (Markowitz et al 2011). Indeed, 'transient' sperm whales in Kaikoura Canyon have been found to exhibit more consistent behavioural effects than 'resident' whales (definition based upon repeated photo-identification of individuals in the area; Gordon et al 1992, Richter et al 2003, Richter et al 2006). The subtle changes in echolocation patterns may therefore indicate tolerance to repeated exposures whereby resident whales are adjusting to the continual presence of whale-watching vessels at surface. Nevertheless, these 'brief' behavioural effects do not necessarily indicate that the transient whales were any less vulnerable to whale-watching than residents (see discussion in Section 1.1; e.g. Beale 2007). Indeed, the transiency of whales may not simply be a feature of individual temperament such as boldness (Dingemanse and Wolf 2010) but also their ability to move to alternative areas (e.g. due to body condition or experience, Section 1.1, Fig. 1.1).

Similarly, sperm whales may have responded to the experimental 1-2 kHz sonar due to a lack of previous experience of the signal, as well as to the high sound levels transmitted from an approaching source (Section 4.4.3). Longer term monitoring would be required to understand whether the response would change over repeated exposures, and whether such behavioural modifications could be considered adaptive rather than maladaptive (Sih 2013) i.e. whether whales would move away from source

levels that could cause physical harm, or spend more time and energy disturbed than necessary for the direct costs that sonar might incur.

The slow life history traits of sperm whales likely contribute to a slow overall population growth rates and comparatively low genetic variability, and subsequently population vulnerability to perturbations (Wade et al 2012, Chiquet et al 2013, Carroll et al 2014). While a likely small proportion of males contribute to the natural-state populations, anthropogenic reductions in male density and/or male quality (e.g. size- and hence sex-biased whaling) could have disproportionate effects due to Allee effects. The Allee effect can be defined as the fitness benefit of conspecific presence (Stephens and Sutherland 1999). Allee effects are of particular concern in endangered species because they can compromise survival of small populations due to depensation (or 'inverse density dependence') (Stephens and Sutherland 1999, Wade et al 2012). Wade et al (2012) suggest that odontocetes may be more vulnerable to such effects than mysticeted because their survival and reproductive success is more dependent on social interactions, including cooperative foraging, defence, parental care and social learning. In terms of a reduced male-to-female ratio, pregnancy rates could fall if female selection remains high for mature or dominant males that are less available, or more indirectly through effects of social disruption on male competition and dispersal (Wade et al 2012). However, there is so far no estimate of a minimum viable population size for male sperm whales. Long-term photo-identification studies of sperm whale groups could be a key to addressing the survival of individuals as a function of group size and male visitation rates. To estimate any population-level consequences of disturbance of sperm whale males at high latitudes, new information on the male migratory pattern is also required. While collaboration of different photo-identification studies have provided first links in the North Atlantic (Steiner et al 2012), specific routes may be described by satellite tags that stay on the animal for months at a time (Mate et al 2007).

6.3 Methods considerations

6.3.1 Multivariate metric vs. state-based approaches

Technological advances in memory, battery power and miniaturisation have made animal-attached tracking devices and sensors available for research across a large range of species and habitats with increasing measurement accuracy and spatio-temporal resolution (Cooke et al 2004, Ropert-Coudert and Wilson 2005, Rutz and Hays 2009, Tomkiewitz et al 2010, Crossin et al 2014). The subsequently energised research in movement ecology and biologging (Naito 2004) has shifted focus from the population-level (Eulerian approach) towards the individual-based (Lagrangian approach) (Nathan et al 2008). Increasing computational power and speed has allowed for post-processing of these large datasets and the development of new computationally intensive analytical tools, such as state-space models (Patterson et al 2008). Consequently, an expanding selection of more multivariate and flexible statistical tools are available to analyse BRS data on marine mammals (e.g. Schick et al 2008, Patterson et al 2008, Schliehe-Diecks et al 2012, McClintock et al 2012) that often deal with correlated measurements, small sample sizes and ordinal and categorical measurements that cannot be analysed using standard Gaussian models.

I explored both multiple pair-wise testing (non-parametric signed rank tests, Chapter 2), generalised linear and generalised additive regression modelling (GLM and GAM, all chapters), state-based and both Bayesian and frequentist classification techniques (Chapters 3-4 and Chapter 5, respectively). Each of these data chapters demonstrated the advantages and pitfalls of the different approaches, highlighting the difficulty of choice of an "optimal statistical approach". Chapter 2 illustrated how multivariate and non-linear regression modelling (GAM) could explain environmental variability in behavioural data, but small sample size meant that this approach was perhaps overly sensitive to finding effects compared to the traditional pair-wise testing. Nevertheless, both approaches suffered from multiple testing: the chance of detecting a false positive result increases with multiple trials (Bland and Altman 1995, Nakagawa 2004). Inherently, this is a problem for defining such a large array of

response variables and touches the very nature of almost all BRS studies: how to best summarise behaviour of interest which is encoded in multiple diverse data time-series. Motivational-behavioural states such as the functional state approach described in Section 1.1 answers this question conceptually. However, even for a well-studied species like the sperm whale, a cost-benefit space may not be feasible to quantify.

The hidden state model developed in Chapter 3 attempted a practical solution to the multiple-metrics problem by defining behaviour in terms of three more simple metrics: biologically informed behaviour state, and two proxy currencies that were measured relative to this behaviour state (buzzing and ODBA). The hidden state model incorporated species-specific expectations of behaviour (echolocation, drifting posture), combined multiple sources of data to estimate biologically interpretable states and parameters (such as descent rate), and allowed modelling of currency proxies within the relevant behavioural contexts. This state-based approach was based upon first principles of searching behaviour (transiting vs. encamped search) and central-place foraging (surface vs. diving) that are transferrable across species, and showed that not all such functional behaviour states need strict definition *a priori* to be estimated. With recent advances in deviance-based selection for Bayesian mixture models (e.g. Plummer 2008), there is increasing scope to incorporate and test a range of disturbance effects as explanatory variables within hidden state models, rather than as a second AIC-based analysis step.

There are three major applications that make hidden process and state-switching models relevant to conservation behaviour. First, response intensity can be specific to context: decision-making mechanisms such as resource selection can be formulated to vary with or be specific to a state (e.g. Getz and Saltz 2008). Behavioural context is increasingly highlighted as the key to understanding the fitness trade-offs of behavioural decision making in response to anthropogenic stimuli (Gill et al 2001, Beale 2007) and within such flexible hierarchical estimation frameworks, could be explicitly modelled by conditioning disturbance effects by behavioural state. Second, changes in internal state (Fig. 1.1) can feed back to behaviour: the probability of switching to or staying in a state can be modelled as a function of subject-specific

explanatory variables, such as body condition or life history stage (e.g. Schliehe-Diecks et al 2012). Third, incorporating state-switching and individual effects allows for more behavioural complexity that may be crucial for scaling up individual movement (Schick et al 2008, Morales et al 2010, Owen-Smith et al 2010) and informing state-dependent life histories (McNamara and Houston 1996).

The challenge and opportunity of the state-based approach is to identify the relevant goals and functional units within behavioural time series at a given spatiotemporal scale. Temporally nested states contain sub-units that may produce multi-modality in sampled behaviour (Nathan et al 2008, Morales et al 2010). For example, movement phases are composed of fine-scale fundamental movement elements (gaits), such as fluking versus gliding (Getz and Saltz 2008, Nathan et al 2008). Scaling up further, motivational states emerge naturally at the three nodes of the Population Consequences of Acoustic Disturbance (PCAD) framework (behavioural change, life function and vital rates) (NRC 2005): behavioural states (e.g. surfacing: harvest oxygen), life history states (e.g. migration: find mates) and population states (e.g. juvenile: survive and maximize body condition for sexual maturity). Promising classifiers include expert-driven signal processing (e.g. Nathan et al 2012), distributions that allow for multimodality (e.g. Yackulic et al 2011) and hidden process models (e.g. Schliehe-Diecks et al 2012) that distinguish the observation and the underlying process explicitly, such as location error from a home-range algorithm (Patterson et al 2008). Hidden process models are increasingly flexible and can incorporate multiple data streams, scales and mechanisms to test hypotheses based on ecological theory by model selection (Schick et al 2008).

6.3.2 Buzzes as a proxy for foraging success in sperm whales

Habitat modelling of acoustic metrics measured remotely from towed hydrophone in Kaikoura Canyon (Section 2.4.3) and clustering of movement metrics and maximum click rate measured onboard DTAGs during buzzing near Lofoten (Chapter 5) supported the idea that sperm whale males employ different foraging strategies in high-latitude foraging grounds (Teloni et al 2008). In Kaikoura, variation in echolocation behaviour

was related to bathymetric features, depth and aspect of the sea floor in particular, and time to first click, suggesting that deeper dives were undertaken in deeper waters. Near Lofoten, the variation was related to both the depth of the whale, as well as the type of prey encounter (i.e. cluster). While there was clear evidence for pressure effects on buzz production (total number of clicks, average and maximum rate, Fig. 5.3), models for within-buzz click rate also indicated that higher acoustic sampling rates were required to capture more mobile prey (Table 5.5). These results corroborate that the sperm whale is a generalist feeder (Whitehead 2003), at least in the high-latitude foraging grounds.

If the nutritional value of prey items varies as much as the movement cost and inter-buzz-interval (by a factor of 10, Table 5.3), then absolute reduction in prey encounter rates alone have little use as a currency proxy, even if capture rates were relatively constant. It is tempting to estimate the average size and calorific value of prey items based on observed buzz rates per dive and assumed mass-specific consumptions (e.g. Drouout et al 2004). Based upon onboard acceleration data, Aoki et al (2012) estimate that sperm whales can expend as much as 5830 kJ on bursts of speed during presumed prey capture attempts. However, if the true distribution of calorie intake is bimodal (or even multimodal), then the average of this distribution has little or no meaning. Transport costs of transiting to great depths may be compensated by consumption of aggregations of more sedentary prey that is easier to detect and capture. Such bimodal or multimodal strategies may be further accentuated by individual variation in body size (oxygen stores).

One advantage of sperm whales being generalist predators and having a relatively large perceptual space for a marine mammal (long detection range for echolocation) is the potential range of available behavioural options in the face of rapid changes in their environment. As illustrated by the differential responses of resident and transient sperm whales to whale-watching however, the available behavioural options likely dependent upon individual context. In the absence of accurate knowledge about these factors, multivariate empirical approaches and before-during-after comparisons of buzz rate (Chapters 2-4) can nevertheless be a useful first step.

6.4 Future research & recommendations

These results and the developed conceptual model (Fig. 1.1) highlight knowledge gaps and research opportunities in conservation behaviour. Further research is needed to assess currencies and state composition within encompassing longer term goals (Nathan et al 2008), life histories (McNamara and Houston 1996) and how such inference may be affected by sampling rate (e.g. Yackulic et al 2011). With this, there is a need to understand and quantify how any short-term cost ("minimum cost", Beale 2007) feeds back to currencies and states over a longer period of time (compensation), both within and across nested states (carry-over effects, O'Connor et al 2014). As well as considering energy and nutrition, also social currencies (e.g. individual associations) are likely to be important links to population dynamics such as density-dependence (Morales et al 2010), particularly when assessing impact of noise on species that communicate acoustically (Laiolo 2010). I envision progress by quantifying these links within a more common conceptual framework, developing process models that can borrow strength across individuals and species, and increasing integration of experimental, observational and multi-scale data.

Specifically, I would like to recommend the following:

I. Whale-watching continues to have subtle effects on echolocation behaviour of sperm whales in a likely critical habitat in Kaikoura. Future monitoring should integrate both land-based and boat-based data, and employ alternative under-water tracking techniques (such as the DTAG) to more conclusively understand effects on foraging behaviour.

II. Tagging procedures are shown to be more related to the presence of tag boat than the device itself, highlighting the importance of consistent protocols for all boat-based focal follow studies of marine mammals.

III. BRS studies focusing on deep-diving marine mammals should assess changes in prey encounter rate given a diving context. Anthropogenic noise might change the relative value of shallow vs. deep foraging by masking but also energetic value of prey may vary with depth.

IV. By incorporating knowledge of behaviour to models we have more power to measure biologically meaningful behavioural responses to anthropogenic disturbance. State-based approaches can be used to control for variability in proxy currencies within states, but future studies should also address any carry-over effects and compensation across states.

6. References

Aoki K, Amano M, Mori K, Kourogi a, Kubodera T, Miyazaki N (2012) Active hunting by deep-diving sperm whales: 3D dive profiles and maneuvers during bursts of speed. Mar Ecol Prog Ser 444:289–301

Beale CM (2007) The behavioral ecology of disturbance responses. Int J Comp Psychol 20:111–120

Bejder L, Samuels a, Whitehead H, Finn H, Allen S (2009) Impact assessment research: use and misuse of habituation, sensitisation and tolerance in describing wildlife responses to anthropogenic stimuli. Mar Ecol Prog Ser 395:177–185

Bland J, Altman D (1995) Multiple significance tests: the Bonferroni method. BMJ 310:1995

Brown J, Laundré J, Gurung M (1999) The ecology of fear: optimal foraging, game theory, and trophic interactions. J Mammal 80:385–399

Carroll G, Hedley S, Bannister J, Ensor P, Harcourt R (2014) No evidence for recovery in the population of sperm whale bulls off Western Australia, 30 years post-whaling. Endanger Species Res 24:33–43

Chiquet R a., Ma B, Ackleh AS, Pal N, Sidorovskaia N (2013) Demographic analysis of sperm whales using matrix population models. Ecol Modell 248:71–79

Cooke SJ, Hinch SG, Wikelski M, Andrews RD, Kuchel LJ, Wolcott TG, Butler PJ (2004) Biotelemetry: a mechanistic approach to ecology. Trends Ecol Evol 19:334–43

Crossin G, Cooke S, Goldbogen J, Phillips R (2014) Tracking fitness in marine vertebrates: current knowledge and opportunities for future research. Mar Ecol Prog Ser 496:1–17

Curé C, Antunes R, Alves A (2013) Responses of male sperm whales (Physeter macrocephalus) to killer whale sounds: implications for anti-predator strategies. Sci Rep 3:1579

Deecke VB, Ford JKB, Slater PJB (2005) The vocal behaviour of mammal-eating killer whales: communicating with costly calls. Anim Behav 69:395–405

Dingemanse NJ, Wolf M (2010) Recent models for adaptive personality differences: a review. Philos Trans R Soc Lond B Biol Sci 365:3947–58

Getz WM, Saltz D (2008) A framework for generating and analyzing movement paths on ecological landscapes. Proc Natl Acad Sci U S A 105:19066–71

Gill J, Norris K, Sutherland W (2001) Why behavioural responses may not reflect the population consequences of human disturbance. Biol Conserv 97:265–268

Laiolo P (2010) The emerging significance of bioacoustics in animal species conservation. Biol Conserv 143:1635–1645

Lusseau D, Bejder L (2007) The long-term consequences of short-term responses to disturbance experiences from whalewatching impact assessment. Int J Comp Psychol 20:228–236

Markowitz T, Richter C, Gordon J (2011) Effects of Tourism on the Behaviour of Sperm whales Inhabiting the Kaikoura Canyon. Report to New Zealand Department of Conservation.

McClintock B, Russell D, Matthiopoulos J, King R (2013) Combining individual animal movement and ancillary biotelemetry data to investigate population-level activity budgets. Ecology 94:838–849

McNamara JM, Houston AI (1996) State-dependent life histories. Nature 380:215–221

Miller PJO, Johnson MP, Madsen PT, Biassoni N, Quero M, Tyack PL (2009) Using at-sea experiments to study the effects of airguns on the foraging behavior of sperm whales in the Gulf of Mexico. Deep Sea Res Part I Oceanogr Res Pap 56:1168–1181

Miller PJ, Kvadsheim PH, Lam, Frans-Peter A Wensveen PJ, Antunes R, Alves AC, Visser F, Kleivane L, Tyack PL, Sivle LD (2012) The Severity of Behavioral Changes Observed During Experimental Exposures of Killer (Orcinus orca), Long-Finned Pilot (Globicephala melas), and Sperm (Physeter macrocephalus) Whales to Naval Sonar. Aquat Mamm 38:362–401

Mate B, Mesecar R, Lagerquist B (2007) The evolution of satellite-monitored radio tags for large whales: One laboratory's experience. Deep Sea Res Part I| 54:224-247

Morales JM, Moorcroft PR, Matthiopoulos J, Frair JL, Kie JG, Powell R a, Merrill EH, Haydon DT (2010) Building the bridge between animal movement and population dynamics. Philos Trans R Soc Lond B Biol Sci 365:2289–301

Naito Y (2004) New steps in bio-logging science. Mem Natl Inst Polar Res:50–57

Nakagawa S (2004) A farewell to Bonferroni: the problems of low statistical power and publication bias. Behav Ecol 15:1044–1045

Nathan R, Getz W (2008) A movement ecology paradigm for unifying organismal movement research. Proc Natl Acad Sci 105:19052–19059

National Research Council (2005) Marine mammal populations and ocean noise: determining when noise causes biologically significant effects. US National Academy of Sciences, Washington DC

Nowacek DP, Thorne LH, Johnston DW, Tyack PL (2007) Responses of cetaceans to anthropogenic noise. Mamm Rev 37:81–115

O'Connor C, Norris D, Crossin G, Cooke S (2014) Biological carryover effects: linking common concepts and mechanisms in ecology and evolution. Ecosphere 5:1–11

Owen-Smith N, Fryxell JM, Merrill EH (2010) Foraging theory upscaled: the behavioural ecology of herbivore movement. Philos Trans R Soc Lond B Biol Sci 365:2267–78

Patterson T a, Thomas L, Wilcox C, Ovaskainen O, Matthiopoulos J (2008) State-space models of individual animal movement. Trends Ecol Evol 23:87–94

Plummer M, Best N, Cowles K (2007) The CODA package. … //www-fis iarc fr/coda

Richter CF, Dawson SM, Slooten E (2003) Sperm whale watching off Kaikoura , New Zealand : effects of current activities on surfacing and vocalisation patterns. Sci Conserv 219

Richter C, Fernandes M, Isojunno S (2011) Effects of vessel traffic on the surface behaviour of individually identified sperm whales foraging in the Kaikoura Canyon. In: Effects of Tourism on the Behaviour of Sperm whales Inhabiting the Kaikoura Canyon.

Ropert-Coudert Y, Wilson RP (2005) Trends and perspectives in animal-attached remote sensing. Front Ecol Environ 3:437–444

Rutz C, Hays GC (2009) New frontiers in biologging science. Biol Lett 5:289–92

Schick RS, Loarie SR, Colchero F, Best BD, Boustany A, Conde D a, Halpin PN, Joppa LN, McClellan CM, Clark JS (2008) Understanding movement data and movement processes: current and emerging directions. Ecol Lett 11:1338–50

Schliehe-Diecks S, Kappeler PM, Langrock R (2012) On the application of mixed hidden Markov models to multiple behavioural time series. Interface Focus 2:180–9

Southall BLB, Bowles AE, Ellison WT (2007) Marine mammal noise-exposure criteria: initial scientific recommendations. Bioacoustics 33

Steiner L, Lamoni L, Acosta Plata M, Jensen S-K, Lettevall E, Gordon J (2012) A link between male sperm whales, Physeter macrocephalus, of the Azores and Norway. J Mar Biol Assoc United Kingdom 92:1751–1756

Stephens P, Sutherland W (1999) Consequences of the Allee effect for behaviour, ecology and conservation. Trends Ecol Evol 14:401–405

Taylor B, Gerrodette T (1993) The uses of statistical power in conservation biology: the vaquita and northern spotted owl. Conserv Biol 7:489–500

Teloni V, Mark JP, Patrick MJO, Peter MT, Johnson MP, Miller PJO, Madsen, Madsen PT (2008) Shallow food for deep divers: Dynamic foraging behavior of male sperm whales in a high latitude habitat. J Exp Mar Bio Ecol 354:119–131

Tomkiewicz SM, Fuller MR, Kie JG, Bates KK (2010) Global positioning system and associated technologies in animal behaviour and ecological research. Philos Trans R Soc Lond B Biol Sci 365:2163–76

Tyack PL (2009) Acoustic playback experiments to study behavioral responses of free-ranging marine animals to anthropogenic sound. Mar Ecol Prog Ser 395:187–200

Wade PR, Reeves RR, Mesnick SL (2012) Social and behavioural factors in cetacean responses to overexploitation: are odontocetes less “resilient” than mysticetes? J Mar Bio 567276:1-15

Williams R, Lusseau D, Hammond PS (2006) Estimating relative energetic costs of human disturbance to killer whales (Orcinus orca). Biol Conserv 133:301–311

Wirsing AJ, Heithaus MR, Frid A, Dill LM (2008) Seascapes of fear: evaluating sublethal predator effects experienced and generated by marine mammals. Mar Mammal Sci 24:1–15

Yackulic CB, Blake S, Deem S, Kock M, Uriarte M (2011) One size does not fit all: flexible models are required to understand animal movement across scales. J Anim Ecol 80:1088–1096

6. Tables

Table 6.1 Severity scoring table, based upon Miller et al (2012) (Table 6A.1).

Score	Avoidance	Change in orientation	Change in locomotion	Change in dive profile	Cessation of foraging	Cessation of resting	Cessation or change in vocal behaviour	Group distribution	Mother-calf separation	Respiration rates	Aggression	Cessation of reproductive behaviour	Other
0													
1		Br											
2		Mo *					Br			Br or Mi			
3		P	Mi	Mi			Mi	Br		Mo			alert behaviour
4	Br		Mo	Mo	**Br**		Mo	Mi					
5	Mi		E	E	**Mi**		P	Mo					
6	Mo				Mo	Mo	E	P	Br or Mi		***	Br	startle response
7	E				P			E	Mo		E / P	Mo	anti-predatory behaviour
8	L								P			P	aversion / sensitization
9	Pr	**									****		stranding event

Br – brief, Mi – minor, Mo – moderate, P – prolonged, E – extended, L – long-term, Pr – predator detection

Bold: additions to Miller et al (2012)

* Or multiple

** Outright flight / panic / stampede

*** Related to noise exposure

**** Attack of conspecifics

Table 6.2 Severity scoring of behavioural responses investigated in this thesis (continued on next page)

Chapter	Exposure type	Exposure duration (min)	Main exposure metrics (sig. metric)	Behavioural changes	Response duration	Severity score
2	Focal-follow vessel 'Titi'	~10-180	Vessel-whale range		-	No detected responses (lack of response perhaps due to lack of control)
2	Whale-watching vessels	~10-60	Presence/ absence	Increased time to first clicks, and increased duration of the following silence	At least 5 min, recovery unknown	2 – brief cessation/ modification of vocal behaviour (possible vertical avoidance / passive listening to vessels)
3	Tagging boat	~5-180	Presence/ absence, time since tag-on, time since tag boat recovery	Increased switching to non-foraging state, and decreased time at surface	Immediate recovery after exposure	2-6 – brief to moderate cessation of feeding (high variability across individuals; vigilance or evasive reaction)
4	Incidental sonar 4.7-5.1 kHz	~300	Presence/ absence, RL, time since, post-exposure window	No robust effects, small increase in buzz rates	-	No detected responses (increased buzz rates likely due to pseudo-replication)
4	Experimental sonar source vessel 'Sverdrup'	35	Presence/ absence, order, time decay, post-exposure window	No effects	-	No detected responses
4	Experimental LFAS 1-2 kHz	35	Presence/ absence, RL, SL, time decay, post-exposure window	Increased switching to and time allocation in silent active state	Recovery 8 min post-exposure	4-6 –moderate cessation of feeding (high variability across individuals; vigilance or evasive reaction)

Table 6.2 continued

Chapter	Exposure type	Exposure duration (min)	Main exposure metrics (sig. metric)	Behavioural changes	Response duration	Severity score
4	MFAS 6-7 kHz	35	Presence/ absence, RL, SL, time decay, post-exposure window	No robust effects	-	No detected responses
4	Killer whale playback boat		Presence/ absence, post-exposure window	No effects	-	No detected responses
4	Killer whale sounds	15	Presence/ absence, post-exposure window	Increased switching to and time allocation in silent active state	Recovery 19 min post-exposure	4-6 – moderate to prolonged cessation of feeding (high variability across individuals; vigilance or evasive reaction)